STUDIES ON AUTOREGRESSIVE TIME SERIES ANALYSIS

Hideaki Sakai

1980

Errata

	erroneous	correct
p.56	(4.16) : $\sum_{t=1}^{N}$	$\sum_{t=1}^{N-\lvert k\rvert}$
p.62	Table 4.2 : 0.014	1.014
p.78	↓12: violates this	violates
p.81	↑8: based on	based on (5.23)
p.95	↑9: $Q(\ _k)$	$Q(\omega_k)$
p.128	(6.79): R(k-i,l)	R(k-i,k)
p.136	↑10: fliter	filter
p.153	↑7: facterization	factorization

STUDIES ON AUTOREGRESSIVE TIME SERIES ANALYSIS

Hideaki Sakai

1980

Acknowledgement

The author wishes to express his sincere thanks to Professor Hidekatsu Tokumaru for his constant guidance and encouragement during the course of this research. The author is also grateful to members of Professor Tokumaru's laboratory for their useful advice. Thanks are particularly due to several people for their assistance in computer programming.

Abstract

This dissertation is concerned with statistical investigations of various modern parametric time series analysis technique. In particular, autoregressive (AR) -type methods are extensively treated and analyzed from an unified view point. That is, through the systematic use of the periodogram, a basic quantity in the frequency domain analysis, their statistical properties are obtained.

After rederiving well known previous results on pure AR processes, a new parameter estimation algorithm based on noisy data is proposed and analyzed. This has applications in image and speech processing.

Then, the periodogram technique is applied to situations where there are randomly or regularly missed observations. This technique is particularly useful since other techniques do not work for these cases. In some cases, the concept of missing observations can be positively utilized to obtain more reliable estimates.

Next, line spectral analyses by AR and Pisarenko methods are considered. The variances of the above two sinusoidal frequency estimation methods are derived. The signal-to-noise (SNR) and the data length play a kind of dual roles in the expressions of the variances at different SNR regions.

Next, some problems on multivariate AR processes are treated. It is shown that statistical properties of the multivariate AR spectral estimate become similar to those of the multivariate Blackman-Tukey method as the fitting order becomes large. A multivariate version of Quenouille's theorem on partial autocorrelations is also obtained. Lastly, a Levinson-type recursive algorithm for Pagano's new method is derived and based on this recursion, a circular lattice structure of the algorithm is shown.

Contents

Chapter 1

Introduction

1.1 A Brief Historical Review

Since Yule (1927) first proposed a method , now called as autoregressive (AR) time series analysis, for analyzing Wölfer's sun-spot numbers, many studies about the AR analysis have been performed because of its simplicity in numerical computation procedures for obtaining the estimates and beautifulness of various theoretical properties.

Here we mention some of the notable researches. The paper of Mann & Wald (1943) is the most fundamental and complete work about the statistical properties of the estimates of the parameters of scalar and multivariate AR processes. Quenouille (1947) introduced the partial autocorrelation coefficients and proposed a statistical test for the determination of autoregression order. However, the hypothesis testing approach to this problem did not succeed in practical situations because of its inherent subjectivity in selecting significant levels. This difficulty is now overcomed by a series of papers of Akaike (1969 a, 1970, 1971, 1973). By his order selecting criterion FPE or AIC, the AR method has been received much more attention and obtaining many good results in various fields of science and engineering.

From a different view point, Burg (1967, 1975) developed the maximum entropy method (MEM) for spectral estimation, now being recognized as a same thing with the AR method. But Burg devised a different estimation algorithm fully utilizing the notion of of partial autocorrelations, or in geophysical terms, reflection coefficients. Independently with Burg's work, Itakura & Saito (1971) also noted a lattice structure of the AR whitening filter which plays a fundamental role in their PARCOR speech analysis-synthesis system. Thus the AR method is not only interesting in theoretical aspects but also is a powerful signal processing technique in actual applications.

1.2 A Preview of This Dissertation

In this dissertation, we investigate further aspects of the AR method from an unified view point, that is , through the systematic use of periodogram, a basic tool in classical time series analysis , for statistical analysis of various modern techniques.

In Chapter 2 , a basic relation between fitting autoregression and periodogram is presented. Using this relation, we re-derive the classical results of Mann & Wald (1943), and the well-known result of Akaike (1969 b) on AR spectral analysis. The technique is also applied to statistical analysis of an AR-MA spectral estimator due to Tokumaru & Takeyasu (1977).

In Chapter 3, we propose a new recursive parameter estimation algorithm for an AR process based on observations corrupted by unknown white noise , and discuss its statistical properties. This algorithm has potential applications to speech and image processing.

In Chapter 4, effects of regularly and randomly missed observations on estimating AR parameters are considered. We can see that the periodogram technique is a powerful tool for this class of problems where most of the conventional techniques break down. In some cases, we can positively utilize missing observations to obtain more reliable estimates.

In Chapter 5, a special , but an important topic in spectral analysis , namely , the estimation of sinusoidal frequencies is treated. We derive the asymptotic variances of the AR and Pisarenko frequency estimators. We particularly note the dependence of the variances on the data length and the signal-to-noise ratio (SNR). They play a kind of dual role in the expression of the variances at different SNR regions.

In Chapter 6 , most of the contents in Chapter 2 are generalized for multivariate AR processes. We show that the normalized reflection coefficient matrices due to Morf *et al.* (1978 a,b) have a desirable property which can be regarded as a generalization of Quenouille's result. The asymptotic order distributions determined by AIC are also derived. The result is a generalization of Shibata (1976). Contrary to our intuition, the probability of selecting the correct order tends to 1 as the number of the variates becomes large. Finally,algorithmic aspects of a new technique due to Pagano (1978) are considerd. We derive a Levinson-type recursive algorithm and show a circular lattice structure of the algorithm which is a multivariate version of the lattice structure of Itakura & Saito (1971). Using this, we devise a Burg-type estimation procedure which guarantees the stability of the estimated filter. The use of the circular lattice filtering to multichannel data compression is suggested.

Chapter 2

Fitting Autoregression and Periodogram

2.1 Introduction

In time series analysis, roughly speaking, mainly two methods have been used. One of them gives the analysis in the frequency domain. It is well known that the quantity called the periodogram plays the fundamental role in that method.(For example, see Jenkins & Watts 1968.) The other gives the analysis in the time domain by which we mean that one postulates some parametric model and the data are fitted to this model by estimating the parameters. Among many models, an autoregressive (AR) process model is preferred because of its simplicity in estimating the parameters. When only the estimation of the autoregressive parameters of a mixed autoregressive-moving average (ARMA) process is of interest, similar simplicity occurs. That is , in the above two cases, one only requires the solution of the Yule-Walker equations. (See , for example, Box & Jenkins 1970.)

The statistical properties of the periodogram and error covariance matrices of the estimates of AR parameters using Yule-Walker equations were discussed and obtained separately. (As for the former, see Jenkins & Watts 1968. As for the later, the classical paper is Mann & Wald ,1943. See also the paper of Baggeroer ,1976 for a different view point.) As far as the author is aware, there are few papers discussing the relation between them.

In this chapter, we present a clear-cut relation connecting them. Then using this formula, the classical results of Mann & Wald (1943) are obtained in Section 2.2 . In Section 2.3, we derive the result of Akaike (1969 b) about the statistical properties of the AR spectral estimator (ARSPE) which will be also used for discussions in Section 5.2 . In Section 2.4, our approach is applied to rederive the error covariance matrix of the estimate of the AR parameters of an ARMA process which was originally obtained by Gersch (1970). Using this, a statistical analysis for an ARMA spectral estimator of Tokumaru & Takeyasu (1977) is performed. From a numerical example, it can be inferred that the variability property of the estimator is rather different from that of the ARSPE.

2.2 Fitting Autoregression and Periodogram

Let us assume for the moment that the time series $\{x_t\}$ under consideration is a zero-mean Gaussian m-th order AR process given by the equation

$$x_t - a_1 x_{t-1} - \dots - a_m x_{t-m} = u_t \tag{2.1}$$

where $\{u_t\}$ is a sequence of white noise with

$$E[u_t] = 0, \quad E[u_t u_s] = \sigma^2 \delta_{t,s} \tag{2.2}$$

To assure the stationarity of (2.1), it is also assumed that the roots of the following equation

$$1 - a_1 z^{-1} - \dots - a_m z^{-m} = 0 \tag{2.3}$$

lie within the unit circle. We denote the autocovariance function of $\{x_t\}$ as $r_k = E[x_{t+k} x_t]$. Define an m×m matrix $\underset{\sim}{R}$, m×1 vectors

$\underset{\sim}{r}$ and $\underset{\sim}{a}$, respectively as follows :

$$(i,k)\text{th element of } \underset{\sim}{R} = (\underset{\sim}{R})_{ik} = r_{i-k}$$

$$\underset{\sim}{r} = (\, r_1, \ldots, r_m \,)^T, \ \underset{\sim}{a} = (\, a_1, \ldots, a_m \,)^T$$

where " T " denotes the transpose operation. As is well known, the Yule-Walker equations hold :

$$\underset{\sim}{R}\,\underset{\sim}{a} = \underset{\sim}{r} \,. \tag{2.3}$$

When a set of data $\{ x_1, \ldots, x_N \}$ is available, one of the most popular estimator for r_k is usually taken as

$$\hat{r}_k = \frac{1}{N} \sum_{t=1}^{N-|k|} x_t \, x_{t+|k|}, \qquad k = 0, \pm 1, \ldots, \pm(N-1), \tag{2.4}$$

Upon substituting these $\hat{r}_k$'s into the r_k's in (2.3), the estimator of $\underset{\sim}{a}$ is given by the solution of

$$\hat{\underset{\sim}{R}}\,\hat{\underset{\sim}{a}} = \hat{\underset{\sim}{r}} \tag{2.5}$$

where $\hat{\underset{\sim}{R}}$, $\hat{\underset{\sim}{a}}$ and $\hat{\underset{\sim}{r}}$ are defined as above. It is well known that as $N \to \infty$, the $\hat{r}_k$'s and $\hat{\underset{\sim}{a}}$ are asymptotically consistent estimators of the r_k's and $\underset{\sim}{a}$, respectively. Define the estimation errors $\Delta\underset{\sim}{a} = \hat{\underset{\sim}{a}} - \underset{\sim}{a}$, $\Delta\underset{\sim}{r} = \hat{\underset{\sim}{r}} - \underset{\sim}{r}$ and $\Delta\underset{\sim}{R} = \hat{\underset{\sim}{R}} - \underset{\sim}{R}$. If N is sufficiently large, these errors can be assumed to be small. By substituting these into (2.5), neglecting the second order terms concerning the errors and noting the relation (2.3), we have

$$\begin{aligned} \underset{\sim}{R}\,\Delta\underset{\sim}{a} &\doteq \Delta\underset{\sim}{r} - (\Delta\underset{\sim}{R})\,\underset{\sim}{a} \\ &= \hat{\underset{\sim}{r}} - \hat{\underset{\sim}{R}}\,\underset{\sim}{a} \,. \end{aligned} \tag{2.6}$$

Let us turn our attention to the periodogram defined by

$$I_N(s) = \frac{1}{2\pi N} \left| \sum_{t=1}^{N} x_t \, e^{-jts} \right|^2, \quad |s| \leqq \pi \tag{2.7}$$

with $j = \sqrt{-1}$. Since (2.7) can be represented in terms of the $\hat{r}_k$'s as

$$I_N(s) = \frac{1}{2\pi} \sum_{k=-N+1}^{N-1} \hat{r}_k e^{-jks}, \tag{2.8}$$

conversely the $\hat{r}_k$'s are represented in terms of the periodogram as

$$\hat{r}_k = \int_{-\pi}^{\pi} I_N(s) e^{jks} ds. \tag{2.9}$$

Hence, the k-th element of (2.6) can be expressed by using (2.9) as follows:

$$\begin{aligned}(\underset{\sim}{R}\Delta\underset{\sim}{a})_k &= \hat{r}_k - \sum_{i=1}^{m} \hat{r}_{k-i} a_i \\ &= \int_{-\pi}^{\pi} B(s) I_N(s) e^{jks} ds\end{aligned} \tag{2.10}$$

where we put

$$B(s) = \sum_{i=0}^{m} (-a_i) e^{-jis} \tag{2.11}$$

with $-a_0 = 1$. On the other hand, it easily follows from (2.8) that for $k > 0$,

$$\begin{aligned}\int_{-\pi}^{\pi} B(s) E[I_N(s)] e^{jks} ds &= \sum_{i=0}^{m} (1 - N^{-1}|k-i|)(-a_i) r_{k-i} \\ &= \eta_k N^{-1}\end{aligned}$$

where we utilize the relation (2.3) and define $\eta_k = \sum_{i=0}^{m} |k-i| a_i r_{k-i}$ for $k > 0$. Therefore, by multiplying the k-th and i-th elements of (2.10) and taking the expectation, the expression of the error covariance matrix is given by

$$\begin{aligned}(\underset{\sim}{R}\, E[\Delta\underset{\sim}{a}\Delta\underset{\sim}{a}^T]\, \underset{\sim}{R}^T)_{k,i} = {} & \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} B(s)B(t) \mathrm{Cov}[I_N(s), I_N(t)] e^{j(ks+it)} ds dt \\ & + \eta_k \eta_i N^{-2}.\end{aligned} \tag{2.12}$$

As will be seen later, the first term in the right hand side of (2.12), which is denoted by $A_{k,i}$ henceforth, is of order N^{-1} so that the second term is of no importance.

Next we examine the estimation error of the residual σ^2 which satisfies the identity

$$\sigma^2 = r_0 - \sum_{k=1}^{m} r_k a_k \,. \qquad (2.13)$$

Hence, the estimator for σ^2 usually takes the form of

$$\hat{\sigma}^2 = \hat{r}_0 - \sum_{k=1}^{m} \hat{r}_k \hat{a}_k \qquad (2.14)$$

where the $\hat{a}_k$'s are given by (2.5). By using the same technique to derive (2.10), the estimation error can be expressed as

$$\begin{aligned}\Delta\sigma^2 &= \Delta r_0 - \sum_{k=1}^{m} (a_k \Delta r_k + r_k \Delta a_k) \\ &= - \Delta \underset{\sim}{a}^T \cdot \underset{\sim}{r} + \int_{-\pi}^{\pi} B(s) I_N(s)\, ds - \sigma^2. \qquad (2.15)\end{aligned}$$

In deriving (2.15), we utilize the relation (2.13) and the fact that $I_N(s) = I_N(-s)$. By virtue of multiplying (2.10) with (2.15) and taking the expectation, we have

$$\begin{aligned}\underset{\sim}{R}\, E[\, \Delta \underset{\sim}{a} \Delta\sigma^2] &\simeq - \underset{\sim}{R}\, E[\, \Delta \underset{\sim}{a} \Delta \underset{\sim}{a}^T]\, \underset{\sim}{r} + \int_{-\pi}^{\pi} B(s) E[\underset{\sim}{R} \Delta \underset{\sim}{a} \cdot I_N(s)]\, ds \\ &\quad - \sigma^2 \underset{\sim}{R} \cdot E[\Delta \underset{\sim}{a}] \,. \qquad (2.16)\end{aligned}$$

Upon noting from (2.12) that $\underset{\sim}{R} \cdot E[\Delta \underset{\sim}{a} \Delta \underset{\sim}{a}^T] = \underset{\sim}{A} \cdot \underset{\sim}{R}^{T-1} = \underset{\sim}{A} \cdot \underset{\sim}{R}$ with $(\underset{\sim}{A})_{k,i} = A_{k,i}$ and $\underset{\sim}{R}^{-1} \underset{\sim}{r} = \underset{\sim}{a}$, the k-th element of the first term of (2.16) becomes $- \sum_{i=1}^{m} A_{k,i} a_i$. The k-th element of the second term is equal to

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} B(s)B(t)\, \mathrm{Cov}[I_N(s), I_N(t)]\, e^{jkt}\, ds \cdot dt + (\sigma^2 + \eta_0 N^{-1}) \eta_k N^{-1}$$

and the k-th element of the third one is $-\sigma^2 \eta_k N^{-1}$. Hence we obtain

$$(\underset{\sim}{R} \cdot E[\, \Delta \underset{\sim}{a} \cdot \Delta\sigma^2])_k = b_k + \eta_0 \eta_k N^{-2} \qquad (2.17)$$

with

$$b_k = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} B(s)B(-s)B(t)\, \mathrm{Cov}[I_N(s), I_N(t)]\, e^{jkt}\, ds \cdot dt \,. \qquad (2.18)$$

Like (2.12), the second term of (2.17) is of no concern.

Denote $\underset{\sim}{R}^{-1} = (q_{k,i})$. Then,

$$\Delta a_k = \sum_{i=1}^{m} q_{k,i} \int_{-\pi}^{\pi} B(t) I_N(t) e^{jit} dt. \quad (2.19)$$

On the other hand, we find that

$$E[(\Delta\sigma^2)^2] = -\sum_{i=1}^{m} r_i E[\Delta\sigma^2 \Delta \underset{\sim}{a}] + \int_{-\pi}^{\pi} B(s) E[\Delta\sigma^2 \cdot I_N(s)]\, ds$$
$$-E[\sigma^2 \Delta\sigma^2]. \quad (2.20)$$

Using (2.17) and the identity

$$\sum_{i=1}^{m} r_i q_{k,i} = a_k , \quad (2.21)$$

the first term of (2.20) is equal to $-\sum_{i=1}^{m} b_i a_i$. The second term can be rewritten by substituting (2.15) into $\Delta\sigma^2$ as

$$-\sum_{k=1}^{m} r_k \int_{-\pi}^{\pi} B(s) \cdot E[\Delta a_k \cdot I_N(s)]\, ds + \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} B(s)B(t) E[I_N(s) I_N(t)] ds \cdot dt$$
$$- \sigma^2 \int_{-\pi}^{\pi} B(s) E[I_N(s)] ds. \quad (2.22)$$

Upon substituting (2.19) into Δa_k in the first term of (2.22) and using (2.21), this first term is given by

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} B(s)B(t)\ \mathrm{Cov}[I_N(s), I_N(t)] \sum_{i=1}^{m} (-a_i) e^{jit}\ ds \cdot dt$$
$$- \sum_{i=1}^{m} a_i (\sigma^2 + n_0 N^{-1}) n_i N^{-1}.$$

Hence, (2.22) finally becomes

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} B(s)B(t)\ \mathrm{Cov}[I_N(s), I_N(t)]\ ds \cdot dt - \sum_{i=1}^{m} a_i (\sigma^2 + n_0 N^{-1}) n_i N^{-1}$$
$$+ (\sigma^2 + n_0 N^{-1})^2 - \sigma^2 (\sigma^2 + n_0 N^{-1}) .$$

Similarly, the third term of (2.20) becomes $\sigma^2 \sum_{i=1}^{m} a_i n_i N^{-1}$ -

$\sigma^2(\ \sigma^2 + \eta_0 N^{-1}) + \sigma^4$. Thereby we finally obtain

$$E[(\ \Delta\sigma^2)^2] \ = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} B(s)B(-s)B(t)B(-t)\mathrm{Cov}[I_N(s),I_N(t)]ds\cdot dt. \tag{2.23}$$

The formulas (2.12), (2.17) and (2.23) give the clear-cut relations between the error covariances in fitting autoregression and the covariance of the periodogram.

On the other hand, if in general $\{\ x_t\ \}$ is a stationary Gaussian time series with the spectral density f(s), the covariance of the periodogram is asymptotically expressed as

$$\mathrm{Cov}[I_N(s),I_N(t)] \ \approx \ f(s)f(t)N^{-2}\{\ F_N(s+t) + F_N(s-t)\} \tag{2.24}$$

where $F_N(\cdot)$ is the Fejér kernel and is defined by

$$F_N(x) \ = \ \frac{\sin^2 \frac{Nx}{2}}{\sin^2 \frac{x}{2}} \quad .$$

(See, for example, Jenkins & Watts 1968, page 250.) Also from the theory of Fejér kernel, it is well known that

$$\int_{-\pi}^{\pi} g(y)\ F_N(x-y)\ dy \ \approx \ 2\pi N\ g(x).$$

Hence, $F_N(\cdot)$ can be approximated by Dirac's delta function as $2\pi N\ \delta(\cdot)$. Thus (2.24) becomes $\mathrm{Cov}[I_N(s),I_N(t)] \doteq 2\pi N^{-1} f(s)f(t)\cdot \{\ \delta(s+t) + \delta(s-t)\}$. Substitution of this into (2.12) gives

$$\begin{aligned} A_{k,i} \ &= 2\pi N^{-1} \int_{-\pi}^{\pi}\{\ B(s)B(-s)f^2(s)\ e^{j(k-i)s} \\ &\quad + B^2(s)f^2(s)\ e^{j(k+i)s}\ \}\ ds \end{aligned} \tag{2.25}$$

The expression (2.12) together with (2.25) can be viewed as the error covariance matrix of the estimate of the optimal linear one-step prediction coefficients of the tapped-delay-line of length m for the general time series. In particular, if $\{x_t\}$ is an m-th order AR process, the power spectra is given by

$$f(s) = \frac{\sigma^2}{2\pi B(s)B(-s)} . \tag{2.26}$$

Thus the first term (2.25) is

$$\sigma^2 N^{-1} \int_{-\pi}^{\pi} \frac{\sigma^2}{2\pi B(s)B(-s)} e^{j(k-i)s}\, ds = \sigma^2 N^{-1} r_{k-i}$$

and the second term is

$$(2\pi N)^{-1}\sigma^4 \int_{-\pi}^{\pi} \frac{1}{B^2(-s)} e^{j(k+i)s}\, ds = \frac{\sigma^2 N^{-1}}{2\pi j} \oint_{|z|=1} \frac{z^{k+i-1}}{B_0^2(z)} dz$$

where $B_0(z) = 1 - a_1 z - \dots - a_m z^m$. Since by the assumption concerning (2.2), the roots of $B_0(z) = 0$ lie outside the unit circle and $k+i-1 > 0$, the integrand of the above complex integral is regular within the unit circle; therefore the integral is zero from Cauchy's theorem. Thus, $\underset{\sim}{A} = \sigma^2 N^{-1}(r_{k-i}) = \sigma^2 N^{-1}\underset{\sim}{R}$. Therefore, the well known result of Mann & Wald (1943)

$$N\, E[\Delta\underset{\sim}{a}\Delta\underset{\sim}{a}^T] \approx \sigma^2 \underset{\sim}{R}^{-1} \tag{2.27}$$

follows. In a similar way, one can show that

$$b_k \quad \frac{\sigma^4 N^{-1}}{\pi j} \oint_{|z|=1} \frac{z^{k-1}}{B_0(z)} dz = 0.$$

Hence,

$$N\, E[\Delta\underset{\sim}{a}\Delta\sigma^2] \approx \underset{\sim}{0}. \tag{2.28}$$

Also from (2.23), we have

$$N\, E[(\Delta\sigma^2)^2] \approx 2\sigma^4 . \tag{2.29}$$

These results are consistent with those obtained in (Box & Jenkins 1970, pages 280-281) by evaluating the Fisher information matrix for the maximum likelihood estimate, if one notes that $\Delta\sigma^2 \approx 2\sigma\Delta\sigma$.

2.3 Autoregressive Spectral Estimation

In this section, we derive the expression for the AR power spectrum estimator originally considered by Akaike (1969 b). Usually, the estimator takes the form of

$$\hat{f}(s) = \frac{\hat{\sigma}^2}{2\,\hat{B}(s)\hat{B}(-s)}, \qquad (2.30)$$

with

$$\hat{B}(s) = 1 - \sum_{i=1}^{m} \hat{a}_i\, e^{-jis}$$

Hence, the estimation error can be approximately represented as

$$\Delta f(s) \approx f(s)\left[\frac{\Delta\sigma^2}{\sigma^2} + \underset{\sim}{H}^T(s)\cdot\Delta\underset{\sim}{a}\right] \qquad (2.31)$$

where we define

$$\underset{\sim}{H}(s) = \frac{\underset{\sim}{E}(s)}{B(s)} + \frac{\underset{\sim}{E}(-s)}{B(-s)} \qquad (2.32)$$

with

$$\underset{\sim}{E}(s) = [\, e^{-js},\, e^{-j2s},\, \ldots,\, e^{-jms}\,]^T.$$

Define the new vector $\underset{\sim}{M}(s)$ by

$$\underset{\sim}{R}^{-1}\underset{\sim}{H}(s) = \underset{\sim}{M}(s) = [\, M_1(s),\, M_2(s), \ldots,\, M_m(s)]^T. \quad (2.33)$$

Then the covariance between the errors at angular frequencies s and t is given by

$$E[\Delta f(s)\Delta f(t)] \approx f(s)f(t)\{\sigma^{-4}E[(\Delta\sigma^2)^2] + \sigma^{-2}[\,\underset{\sim}{M}^T(s) + \underset{\sim}{M}^T(t)]$$
$$\times\underset{\sim}{R}\cdot E[\Delta\underset{\sim}{a}\Delta\sigma^2] + \underset{\sim}{M}^T(s)\underset{\sim}{R}\cdot E[\Delta\underset{\sim}{a}\Delta\underset{\sim}{a}^T]\underset{\sim}{R}^T\underset{\sim}{M}(t)\}\ . \quad (2.34)$$

By substituting (2.12), (2.17) and (2.23) into (2.34) and defining

$$K(s,t) = \sum_{k=1}^{m} M_k(s)\, e^{jkt}, \qquad (2.35)$$

we obtain the desired formula

$$E[\Delta f(s)\Delta f(t)] \doteq f(s)f(t)\cdot\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} B(\mu)B(\nu)\mathrm{Cov}[I_N(\mu),I_N(\nu)]$$
$$[K(s,\mu)+\sigma^{-2}B(-\mu)][K(t,\nu)+\sigma^{-2}B(-\nu)]\,d\mu\cdot d\nu\ . \qquad (2.36)$$

When $\{x_t\}$ is Gaussian, by the same calculations leading to the results (2.27), (2.28) and (2.29) one can show that

$$N\cdot E\left[\frac{\Delta f(s)}{f(s)}\cdot\frac{\Delta f(t)}{f(t)}\right] \approx 2+\sum_{k=1}^{m}\sum_{i=1}^{m} M_k(s)M_i(t)r_{k-i}. \qquad (2.37)$$

The second term of (2.37) is further simplified by using (2.33) to $\underset{\sim}{H}^T(s)\underset{\sim}{R}^{-1}\underset{\sim}{H}(t)$. Using the matrix factorization for $\underset{\sim}{R}^{-1}$ described in (3.7) of Akaike (1969 b), we obtain the same formula originally given by Akaike. At this point, the Akaike's result is more general than ours since the former is free from the Gaussian assumption. However, this drawback is remedied as follows. If one drops this Gaussian assumption, one must add a term of order N^{-1} to the right hand side of (2.24). This term can be obtained by the following argument. Let the time series $\{x_t\}$ be generated by passing the innovation sequence $\{u_t\}$ whose variance and fourth cumulant are σ^2 and κ_4, respectively, through the linear filter with the transfer function G(s). Then, by the formula (6.3.15) in Jenkins & Watts (1968, page 238) and the argument developed in page 250 of that reference, the extra term can be expressed as $N^{-1}\kappa_4|G(s)|^2\sigma^2 \times|G(t)|^2\sigma^2/\sigma^4 = N^{-1}\kappa_4 f(s)f(t)/\sigma^4$ where f(s) is the spectra of $\{x_t\}$. That is,

$$\mathrm{Cov}[I_N(s),I_N(t)] \doteq 2\pi N^{-1}f(s)f(t)[\delta(s+t)+\delta(s-t)]$$
$$+\ N^{-1}f(s)f(t)\kappa_4\sigma^{-4}. \qquad (2.38)$$

It can easily be shown that in the presence of this extra term the results (2.27) and (2.28) are unchanged while in the right hand sides of (2.29) and (2.37) one must add κ_4 and κ_4/σ^4, respectively. Thus we can obtain the results of Mann & Wald (1943) and Akaike (1969 b) without the Gaussian assumption.

An important property of (2.37) is that if we take the fitted autoregression order M much greater than the true order m, we can easily show that

$$N\,E[(\frac{\Delta f(s)}{f(s)})^2] \approx \begin{cases} 2M & \text{for } s \neq 0, \pi \\ 4M & \text{for } s = 0, \pi . \end{cases} \qquad (2.39)$$

The proof of (2.39) is deferred to Chapter 6 where a multivariate version of (2.39) will be stated and proved. The meaning of (2.39) is that the statistical variabilities of AR spectral estimation are quite stable over all range of the frequencies. Berk (1974) has shown that the property (2.39) also holds for a certain class of infinite order AR processes with infinitely increasing fitted autoregression order M .

2.4 Statistical Analysis of an ARMA Spectral Estimator

2.4.1 Introduction & Summary

In this section, we first rederive the error covariance matrix of the estimate of the AR parameters of an ARMA process which was originally obtained by Gersch (1970). Since several authors have recently proposed a spectral estimator for ARMA processes which involves only the estimates for AR parameters and some of the auto-covariances of the process without solving the nonlinear equations for MA parameters (Tokumaru & Takeyasu 1977, Kinkel *et al.* 1978, and Kaveh 1978), the asymptotic variances of the above estimator are derived by using the periodogram technique.

However, the resulting formula is rahter complicated so that numerical calculations are performed for a simple example. The result indicates that the variabilities strongly depend on frequencies. This property is quite different that of the well known AR spectral estimator (2.39).

2.4.2 Another Derivation of Gersch's Result

We first consider the error covariance matrix for the Yule-Walker estimate of AR parameters of a Gaussian ARMA(m,n) process generated by

$$x_t - a_1 x_{t-1} - \dots - a_m x_{t-m} = u_t - b_1 u_{t-1} - \dots - b_n u_{t-n} \tag{2.40}$$

with $E[u_t] = 0$, $E[u_t u_s] = \sigma^2 \delta_{t,s}$. It is also assumed that the roots of $B_0(z) = 1 - a_1 z - \dots - a_m z^m = 0$ lie outside the unit circle. Obviously, the Yule-Walker equations

$$\bar{\underset{\sim}{R}}_n \underset{\sim}{a} = \bar{\underset{\sim}{r}}_n$$

hold where $(\bar{\underset{\sim}{R}}_n)_{k,i} = r_{n+k-i}$, $(\bar{\underset{\sim}{r}}_n)_i = r_{n+i}$. Hence, as above, the covariance matrix is given by

$$\begin{aligned}(\bar{\underset{\sim}{R}}_n E[\Delta\underset{\sim}{a}\Delta\underset{\sim}{a}] \bar{\underset{\sim}{R}}_n^T)_{k,i} &\approx 2\pi N^{-1} \int_{-\pi}^{\pi} B^2(s) f^2(s) e^{j(k+i+2n)s} ds + \\ &\quad 2\pi N^{-1} \int_{-\pi}^{\pi} B(s)B(-s) f^2(s) e^{j(k-i)s} ds \end{aligned} \tag{2.41}$$

where f(s) is the spectra of (2.40) and is given by

$$f(s) = \frac{C(s)C(-s)}{2\pi B(s)B(-s)} \sigma^2$$

with $C(s) = \sum_{i=0}^{n} (-b_i)\exp(-jis)$ $(-b_0 = 1)$. By rewriting the first term of (2.41) as a complex integral, one can easily show from the assumption concerning the a_i's that this integral is equal to zero.

The second term is

$$\sigma^2 N^{-1} \sum_{q=0}^{n} \sum_{q'=0}^{n} (-b_q)(-b_{q'}) r_{k-i-q+q'} .$$

By defining

$$\gamma_k = \sum_{i=0}^{n-k} (-b_i)(-b_{i+k}), \quad k= 0, 1, \ldots, n-1,$$

the right hand side of (2.41) is written in a matrix form as

$$\bar{\underset{\sim}{A}} = \sigma^2 N^{-1} [\gamma_0 \bar{\underset{\sim}{R}}_0 + \sum_{i=1}^{n} \gamma_i (\bar{\underset{\sim}{R}}_i + \bar{\underset{\sim}{R}}_i^T)] \quad (2.42)$$

which coincides with the result of Gersch (1970).

2.4.3 A Spectral Estimator and Its Statistical Analysis

In this subsection, we restrict our attention to the case $n = m-1$ in (2.40) since such an ARMA(m,m-1) process arises as a result of sampling a physically realizable continuous-time linear stochastic process (Kinkel *et al.* 1978). The Yule-Walker equations are

$$r_{m+i} - a_1 r_{m+i-1} - \ldots - a_m r_i = 0, \; i = 0, 1, 2, \ldots (2.43)$$

By multiplying both sides of (2.43) by $\exp(-jsi)$ and summing them from $i = 0$ to ∞, it follows that

$$\sum_{k=0}^{m-1} (-a_k) \, e^{js(m-k)} \{ Q(s) - \sum_{i=0}^{m-k-1} r_i \, e^{-jsi} \} = 0 \qquad (2.44)$$

where we define

$$Q(s) \triangleq \sum_{i=0}^{\infty} r_i \, e^{-jsi} \quad \text{and } a_0 = -1. \qquad (2.45)$$

Hence from (2.44), $Q(s)$ is written as

$$Q(s) = \frac{1}{\bar{B}(s)} \sum_{k=0}^{m-1} \sum_{i=0}^{m-k-1} (-a_k) r_i \, e^{js(m-k-i)} \quad (2.46)$$

with

$$\bar{B}(s) = \sum_{i=0}^{m} (-a_i) \, e^{j(m-i)s} \, . \quad (2.47)$$

Hence, the spectra of tne process can be expressed as

$$f(s) = \frac{1}{2\pi} [\, Q(s) + Q^*(s) - r_0 \,] \quad (2.48)$$

where " $*$ " denotes the complex conjugate operation. Note that these formulas do not contain MA parameters so that the estimation procedure is quite easy. In fact, combining the first m equations in (2.43), we have

$$\bar{\underset{\sim}{R}} \, \underset{\sim}{a} = \bar{\underset{\sim}{r}} \quad (2.49)$$

where for abbreviation we put $\bar{\underset{\sim}{R}}_{m-1} = \bar{\underset{\sim}{R}}$ and $\bar{\underset{\sim}{r}}_{m-1} = \bar{\underset{\sim}{r}}$. Then, the estimation error is expressed as

$$(\bar{\underset{\sim}{R}} \Delta \underset{\sim}{a})_k \approx \int_{-\pi}^{\pi} \bar{B}(s) I_N(s) \, e^{j(-1+k)s} \, ds \quad (2.50).$$

Using the classical result of Mann & Wald (1944), $\Delta Q(s) = \hat{Q}(s) - Q(s)$ is asymptotically equal to

$$\Delta Q(s) \quad \frac{\partial Q(s)}{\partial \underset{\sim}{r}} (\hat{\underset{\sim}{r}} - \underset{\sim}{r}) + \frac{\partial Q(s)}{\partial \underset{\sim}{a}} (\hat{\underset{\sim}{a}} - \underset{\sim}{a})$$

$$= - \frac{1}{\bar{B}(s)} \sum_{k=0}^{m-1} \{ \sum_{i=0}^{m-k-1} (\Delta a_k \cdot r_i + a_k \hat{r}_i - a_k r_i)$$

$$\times e^{js(m-k-i)} \} + \frac{Q(s)}{\bar{B}(s)} \sum_{i=0}^{m} \Delta a_i \, e^{js(m-i)} \quad (2.51)$$

where $\underset{\sim}{r} = (r_1, \ldots, r_m)^T$ and $\hat{r}_k$' are given by (2.4) with $\Delta a_0 = 0$. Substituting (2.9) into $\hat{r}_i$'s in (2.51), we obtain

$$\Delta Q(s) \approx \bar{\underset{\sim}{H}}^T(s) \Delta \underset{\sim}{a} + \int_{-\pi}^{\pi} \bar{F}(s,t) I_N(t) dt - Q(s) \quad (2.52)$$

where we define $\underset{\sim}{\bar{H}}^T(s) = (\bar{H}_1(s),\dots,\bar{H}_m(s))$ with

$$\bar{H}_k(s) = \frac{1}{\bar{B}(s)} \{ \sum_{i=0}^{m-k-1} r_i \, e^{js(m-k-i)}(1-\delta_{m,k}) - Q(s)e^{js(m-k)} \} \tag{2.53}$$

$$\bar{F}(s,t) = \frac{1}{\bar{B}(t)} \sum_{k=0}^{m-1} e^{jks} \sum_{i=0}^{m-k-1} (-a_i) \, e^{jt(m-k-i)}$$

$$= \sum_{k=0}^{m-1} \bar{J}_k(t) \, e^{jks}. \tag{2.54}$$

To express the first term of (2.52) in terms of the periodogram , the new vector $\underset{\sim}{\bar{M}}(s)$ is defined by $\underset{\sim}{\bar{H}}^T(s)\underset{\sim}{\bar{R}}^{-1} = \underset{\sim}{\bar{M}}^T(s) = (\bar{M}_1(s), \dots, \bar{M}_m(s))^T$. Then from (2.50) it becomes

$$- \int_{-\pi}^{\pi} \bar{B}(t)K(t,s)I_N(t) \, dt \tag{2.55}$$

with

$$\bar{K}(t,s) = \sum_{k=1}^{m} \bar{M}_k(s) \, e^{j(-1+k)t}. \tag{2.56}$$

Thus, from (2.48) the estimation error of $\hat{f}(s)$ is given by

$$\Delta f(s) \;:\; \frac{1}{2\pi} [\, \Delta Q(s) + \Delta Q^*(s) - \hat{r}_0 + r_0 \,]$$

$$= \int_{-\pi}^{\pi} \bar{G}(t,s)I_N(t) \, dt - f(s) \tag{2.57}$$

with

$$2\pi \, \bar{G}(t,s) = -\bar{B}(t)\bar{K}(t,s)+ \bar{F}(t,s) + [-\bar{B}(t)\bar{K}(t,s)+\bar{F}(t,s)]^*-1 \tag{2.58}$$

By the same argument in Appendix 5.1, we can show that

$$\int_{-\pi}^{\pi} \bar{G}(t,s)f(t) \, dt = f(s). \tag{2.59}$$

This shows that $f(s)$ is one of the eigenfunction of the kernel

$\bar{K}(t,s)$. From (2.57), (2.59) and noting $E[I_N(s)] = f(s) + O(N^{-1})$, the error covariance between angular frequencies ω and η is

$$E[\ \Delta f(\omega)\Delta f(\eta)\ ] \approx \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \bar{G}(s,\omega)\bar{G}(t,\eta)\mathrm{Cov}[I_N(s),I_N(t)]\ dsdt\ . \tag{2.60}$$

On the other hand, the covariance of $I_N(s)$ is given by (2.38). Thus we have

$$N\ E[\ \frac{\Delta f(\omega)}{f(\omega)}\cdot\frac{\Delta f(\eta)}{f(\eta)}\ ] \approx \kappa_4\ \sigma^{-4} + \frac{2\pi}{f(\omega)f(\eta)} \int_{-\pi}^{\pi}[\bar{G}(s,\omega)\bar{G}(s,\eta) + G(s,\omega)G(-s,\eta)]\ f^2(s)\ ds. \tag{2.61}$$

From (2.47), (2.54), (2.56), and (2.58), the kernel $\bar{G}(s,\omega)$ can be written as

$$\bar{G}(s,\omega) = \sum_{k=-2m+1}^{2m-1} \bar{g}_k(\omega)\ e^{jks} \quad (\ g_{-k}(\omega) = g_k^*(\omega)\ ) \tag{2.62}$$

with appropriate functions $g_k(\omega)$'s. Substitution of (2.62) and $f(s) = \sum_{i=-\infty}^{\infty} r_i\ \exp(-jis)/2\pi$ into (2.61) gives the evaluation of (2.61) as

$$\kappa_4\ \sigma^{-4} + (f(\omega)f(\eta))^{-1} \sum_{k=-2m+1}^{2m-1}\ \sum_{k'=-2m+1}^{2m-1}\ \sum_{i=-\infty}^{\infty} \bar{g}_k(\omega)\bar{g}_{k'}(\eta) \times r_i\ [\ r_{k+k'-i} + r_{k-k'-i}\ ]. \tag{2.63}$$

Regrettably, we could not obtain a simpler and more interpretative formula than (2.63), but by numerical calculations the statistical properties can be examined.

2.4.4 An Example and Concluding Remarks

To examine the property of (2.63), numerical calculations were performed for the following ARMA(2,1) process:

$$x_t - a_1 x_{t-1} - \dots - a_2 x_{t-2} = u_t - b_1 u_{t-1}.$$

In this case,

$$\bar{g}_3(\omega) = -\bar{M}_2(\omega),$$

$$\bar{g}_2(\omega) = -\bar{M}_1(\omega) + a_1\bar{M}_2(\omega),$$

$$\bar{g}_1(\omega) = a_1\bar{M}_1(\omega) + a_2\bar{M}_2(\omega) + \bar{J}_1(\omega)$$

and

$$\bar{g}_0(\omega) = a_2[\ \bar{M}_1(\omega) + \bar{M}_1^*(\omega)\ ] -1 + \bar{J}_0(\omega) + \bar{J}_0^*(\omega)$$

where

$$\bar{M}_1(\omega) = [\ \bar{H}_1(\omega)r_1 - \bar{H}_2(\omega)r_2]/(r_1^2 - r_0 r_2),$$

$$\bar{M}_2(\omega) = [-\bar{H}_1(\omega)r_0 + \bar{H}_2(\omega)r_1]/(r_1^2 - r_0 r_2)$$

with

$$\bar{H}_1(\omega) = [\ r_0 - Q(\omega)\ ]\ e^{j\omega}/\bar{B}(\omega),$$

$$\bar{H}_2(\omega) = -Q(\omega)/\bar{B}(\omega)$$

and

$$\bar{J}_1(\omega) = e^{j\omega}/\bar{B}(\omega).$$

The r_i's are given by $r_i = a_1 r_{i-1} + a_2 r_{i-2}$ $(i \geqq 2)$ with starting values satisfying

$$(1-a_2^2)r_0 - a_1(1+a_2)r_1 = (1+b_1^2 - b_1 a_1)\ \sigma^2$$

$$-a_1 r_0 + \ (1-a_2)r_1 = -b_1\ \sigma^2.$$

For $a_1 = 0.75$, $a_2 = -0.5$, $b_1 = -0.5$, $\sigma^2 = 1$, and $\kappa_4 = 0$ (Gaussian distribution), (2.63) was computed by truncating the infinite summation at ± 40 and setting $\omega - \eta$. Table 2.1 shows these numerical results together with the values of f(s) and the empirical values of (2.63) which were obtained by averaging the squares of relative spectral estimation errors over 300 different data sets each of 1500 length. From this table it can be seen that the variabilities of $\hat{f}(s)$, that is, $E[(\Delta f(s)/f(s))^2]$ strongly depend on frequencies. In particular, at frequencies where the values of $f(s)$ are extremely small, the variabilities become quite large. The autoregressive spectral estimator does not possess this property, since the variabilities are expressed as (2.39).

ω $k\pi/25$	Spectrum	Values of (2.63)	Empirical Values
k = 0	0.6366	9.9506	9.8445
2	0.6930	9.2785	9.1495
4	0.8985	7.6608	7.5994
6	1.3396	8.9746	9.5862
8	1.5693	16.308	14.475
10	0.8051	9.0822	9.9216
12	0.3077	9.2159	9.5496
14	0.1297	8.3152	8.8510
16	0.0615	5.9255	6.5007
18	0.0319	3.9507	4.6919
20	0.0178	7.0696	8.7905
22	0.0110	23.701	28.334
24	0.0082	50.390	59.153

Table 2.1 The numerical values of theoretical and experimental variabilities.

Another important problem which is not treated here is the determination of the order m from the data samples. At present, we feel that Akaike's canonical correlation analysis (Akaike 1976) is perhaps a most suitable, though rather complicated, tool for this purpose.

2.5 Conclusion

In this chapter, we have presented a basic relation between fitting autoregression and periodogram. As can be seen from the covariance property (3.28) of the periodogram $I_N(s)$, we can regard $I_N(s)$ as a " white noise " process by taking the frequency s as the usual time parameter. Then the relations (2.10), (2.23) , (2.50), and (2.57) are nothing but the orthogonal spectral decompositions of the estimators, just as the well known spectral expression of stationary processes (for example, Brillinger 1975, Chapter 3). This property permits immediate calculations of covariances of various estimators, once we can asymptotically express them as linear functionals of the periodogram.

We have applied this technique to rederive the results of Mann & Wald, Akaike, and Gersch. We have also obtained the statistical variabilities of an ARMA spectral estimator which essentially belongs to AR techniques.

Appendix 2.1

Here we give a useful expression of $\Delta\sigma^2$ for later discutions.

From (2.15) and (2.10) we have

$$\Delta\sigma^2 \approx -\Delta \underset{\sim}{a}^T \underset{\sim}{r} + \hat{r}_0 - r_0 - \underset{\sim}{a}^T(\hat{\underset{\sim}{r}} - \underset{\sim}{r})$$

$$= -(\underset{\sim}{R}\Delta\underset{\sim}{a})^T \underset{\sim}{R}^{-1} \underset{\sim}{r} + \int_{-\pi}^{\pi} B(s) I_N(s)\, ds - \sigma^2$$

$$= \int_{-\pi}^{\pi} B(s)B(-s) I_N(s)\, ds - \sigma^2. \qquad \text{(A-2.1)}$$

Chapter 3

Recursive Parameter Estimation of an AR Process Disturbed by White Noise

3.1 Introduction

The problem of estimating the autoregressive (AR) process parameters based on the data corrupted by unkown white noise has been studied by many authors. This problem was first treated by Walker (1960). Later, Parzen (1967) proposed some methods to estimate the process parameters and signal-to-noise ratio (SNR) by means of spectral density and third order correlation. Kashyap (1970) discussed the maximum likelihood method for more general vector mixed autoregressive-moving average model with additional white noise. Further, Pagano (1974) presented the non-linear regression method which is asymptotically consistent with the maximum likelihood method and proved its properties. Tong (1975) discussed the method to determine the order of AR process using Akaike's information criterion, where the parameters are estimated by the maximum likelihood method with constrained conditions. In all the above cases, estimation methods are off-line, or non-recursive, but in many practical situations, process parameters may fluctuate faintly. So it may be desirable that the estimation method is updated with the data, or recursive.

In this chapter, we mainly examine two methods. The one is an on-line version of the solutions of the well known Yule-Walker

(Y-W) equations and generally is used as the preliminary estimate for the maximum likelihood method, where in this case the statistical fluctuation of this estimate may be large because higher lag autocovariance estimates must be used. Although this method does not require much numerical calculations, its efficiency is relatively low (Walker 1960). The other is based on the direct application of the least-squares method to the observed data. obviously in this case, the resulting estimate is biased, so that we must compensate this bias. Hence we apply the modified least-squares (MLS) method previously proposed by Sagara & Wada (1977) for the parameter estimation problem of a transfer function based on known input data and output data corrupted by white noise. It is intuitively expected that the fluctuation of this estimate is relatively smaller than that of the Yule-Walker estimate because lower lag autocovariance estimates are used. For comparison, the bootstrap (BS) method described by Mayne (1967) and Rowe (1970) is also applied to this problem.

Using the periodogram technique in Chapter 2, we first derive the asymptotic error covariance matrix of the Y-W estimate. Although Walker (1960) already obtained it for the first order AR process, our result is more general. Next, we derive that of the MLS estimator. Since for the first order AR process, both estimators are seen to be identical, we compare their performances for the second order AR process. To see the appropriateness of the asymptotic analysis, the empirical error covariances obtained by simulations are also presented along with the corresponding Cramér-Rao bound

(Pagano 1974). Lastly, sample runs of the two recursive estimators are shown to indicate their long-range behaviours also. From these numerical experiments, it can be seen that the results of the simulations and the asymptotic analysis show consistencies surprisingly well and the MLS method is best among the three.

3.2 The Yule-Walker Estimate

Let us assume as in Chapter 2 that the time series $\{x_t\}$ under consideration is a zero-mean Gaussian m-th order AR process given by the equation

$$x_t - a_1 x_{t-1} - \dots - a_m x_{t-m} = u_t \tag{3.1}$$

where $\{u_t\}$ is a sequence of white noise with $E[u_t] = 0$, $E[u_t u_s] = \sigma^2 \delta_{t,s}$. We consider the situation where $\{x_t\}$ is not directly observed but the noisy version of it is available. Denote the noisy observation sequence by $\{y_t\}$ with

$$y_t = x_t + n_t \tag{3.2}$$

where $\{n_t\}$ is a Gaussian white noise sequence uncorrelated with $\{x_t\}$, that is

$$E[n_t] = 0, \quad E[n_t n_s] = \sigma_n^2 \delta_{t,s}, \quad E[n_t u_s] = 0.$$

We wish to estimate parameters $a_1, a_2, \dots, a_m, \sigma^2$, and σ_n^2 recursively when we have N successive samples $y_1, y_2, \dots, y_N$. From (3.1) and (3.2) we have

$$y_t - a_1 y_{t-1} - \dots - a_m y_{t-m} = u_t + n_t - a_1 n_{t-1} - \dots - a_m n_{t-m} \tag{3.3}$$

Then (3.3) may be rewritten in the following form

$$y_t - \underset{\sim}{Y}_{t-1}^T \underset{\sim}{a} = \nu_t \tag{3.4}$$

where

$$\underset{\sim}{Y}_{t-1} = [\, y_{t-1}, y_{t-2}, \ldots, y_{t-m} \,]^T$$

$$\underset{\sim}{a} = [\, a_1, a_2, \ldots, a_m \,]^T$$

We denote $E[y_t y_{t+k}] = r_k$, $E[x_t x_{t+k}] = r'_k$ so $r_k = r'_k + \delta_{k,0}$.

Now we present a very simple and easy estimation method based on the Yule-Walker equations. Premultiplying both sides of (3.4) by $\underset{\sim}{Y}_{t-m-1}$, taking the expectation, and using $E[\underset{\sim}{Y}_{t-m-1}\nu_t] = \underset{\sim}{0}$, it follows that

$$E[\underset{\sim}{Y}_{t-m-1}\underset{\sim}{Y}_{t-1}^T]\, \underset{\sim}{a} = E[\underset{\sim}{Y}_{t-m-1} y_t] \tag{3.5}$$

Thus it is obtained from (3.5) that

$$\hat{\underset{\sim}{a}}_N = [\, \sum_{t=1}^{N} \underset{\sim}{Y}_{t-m-1}\underset{\sim}{Y}_{t-1}^T \,]^{-1} [\, \sum_{t=1}^{N} \underset{\sim}{Y}_{t-m-1} y_t \,]. \tag{3.6}$$

By putting $\underset{\sim}{V}_N = [\, \sum_{t=1}^{N} \underset{\sim}{Y}_{t-m-1}\underset{\sim}{Y}_{t-1}^T \,]^{-1}$, the recursive formula for $\underset{\sim}{V}_N$ is given by

$$\begin{aligned} \underset{\sim}{V}_{N+1} &= \underset{\sim}{V}_N - \underset{\sim}{V}_N \underset{\sim}{Y}_{N-m} W_N^{-1} \underset{\sim}{Y}_N^T \underset{\sim}{V}_N \\ W_N &= 1 + \underset{\sim}{Y}_N^T \underset{\sim}{V}_N \underset{\sim}{Y}_{N-m} \end{aligned} \tag{3.7}$$

Also it is easy to see that the recursive formula for $\hat{\underset{\sim}{a}}_N$ in (3.6) is given by

$$\hat{\underset{\sim}{a}}_{N+1} = \hat{\underset{\sim}{a}}_N + \underset{\sim}{V}_N \underset{\sim}{Y}_{N-m} (y_{N+1} - \underset{\sim}{Y}_N^T \hat{\underset{\sim}{a}}_N) W_N^{-1} \tag{3.8}$$

Although the recursive formulas for $(\hat{\sigma}^2)_N$ and $(\hat{\sigma}_n^2)_N$ can be easily obtained, these are omitted here. The bootstrap estimator is obtained by replacing $\underset{\sim}{Y}_{t-m-1}$ by $\hat{\underset{\sim}{Y}}_{t-1}$ which is a kind of estimator for $\underset{\sim}{Y}_{t-1}$ and is generated from $\underset{\sim}{Y}_{t-m-1}$ by the algorithm due to Rowe (1970).

Now we derive the error covariance matrix of the above $\hat{\underset{\sim}{a}}_N$. We rewrite (3.5) as

$$\tilde{\underset{\sim}{R}}\, \underset{\sim}{a} = \tilde{\underset{\sim}{r}} \tag{3.9}$$

where $(\tilde{\underset{\sim}{R}})_{k,i} = r_{m+k-i}$, $(\tilde{\underset{\sim}{r}})_i = r_{m+i}$. Substituting $\hat{r}_k$ defined by

$$\hat{r}_k - \frac{1}{N} \sum_{i=1}^{N-|k|} y_i y_{i+|k|} \tag{3.10}$$

into r_k in (3.9) yields the Yule-Walker estimate $\hat{\underset{\sim}{a}}$ which is asymptotically equal to $\hat{\underset{\sim}{a}}_N$ in (3.6). By the same argument developed in Chapter 2, it follows that the estimation error $\Delta\underset{\sim}{a} = \hat{\underset{\sim}{a}} - \underset{\sim}{a}$ satisfies

$$(\tilde{\underset{\sim}{R}}\Delta\underset{\sim}{a})_k \approx \int_{-\pi}^{\pi} B(s)\, I_N(s)\, e^{j(k+m)}\, ds$$

where $I_N(s)$ is the periodogram for the process $\{y_t\}$ and $B(s)$ is defined by (2.11). Hence, corresponding to (2.12) we have

$$\begin{aligned}(\tilde{\underset{\sim}{R}}\, E[\Delta\underset{\sim}{a}\Delta\underset{\sim}{a}^T]\, \tilde{\underset{\sim}{R}}^T)_{k,i} &\approx \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} B(s)B(t)\mathrm{Cov}[I_N(s), I_N(t)] \\ &\quad \times e^{j[(k+m)s+(i+m)t]}\, ds\cdot dt = \tilde{A}_{k,i}.\end{aligned} \tag{3.11}$$

By the same approximate calculations as in Chapter 2 and noting that the spectra of $\{y_t\}$ is given by

$$f_y(s) \quad \frac{1}{2\pi}\left[\sigma_n^2 + \frac{\sigma^2}{B(s)B(-s)}\right]$$

, (3.11) becomes

$$\frac{1}{2\pi N} \int_{-\pi}^{\pi} B^2(s)\left[\sigma_n^2 + \frac{\sigma^2}{B(s)B(-s)}\right]^2 e^{j(k+i+2m)} ds$$

$$+ \frac{1}{2\pi N} \cdot \int_{-\pi}^{\pi} B(s)B(-s)\left[\sigma_n^2 + \frac{\sigma^2}{B(s)B(-s)} \right]^2 e^{j(k-i)s}\, ds. \quad (3.12)$$

If we rewrite the first term of (3.12) as a complex integral, it can easily be seen that this integral is equal to zero. Hence, we finally obtain

$$A_{k,i} = N^{-1}(e_{k,i}\, \sigma_n^4 + 2\sigma_n^2\sigma^2\delta_{k,i} + \sigma^2 r_{k-i}). \quad (3.13)$$

with

$$e_{k,i} = \sum_{q-q'=k-i} (-a_q)(-a_{q'}); \; (q,q'=0,1,\dots,m).$$

As an example, let $\{x_t\}$ be a first-order AR process with $r'_k = \sigma^2 a^{|k|}/(1-a^2)$. Then, (3.11) together with (3.13) reduces to

$$N\cdot E[(\Delta a)^2] \approx \frac{1-a^2}{a^2}[1+2\lambda(1-a^2)+\lambda^2(1-a^4)] \quad (3.14)$$

with $\lambda = \sigma_n^2/\sigma^2$. This result, of course, agrees with the one obtained by Walker (1960).

3.3 The Modified Least-Squares Estimate

By applying the least-squares to (3.4) directly, we obtain

$$\left[\sum_{t=1}^{N} \underline{Y}_{t-1}\underline{Y}_{t-1}^T \right] \hat{\underline{a}}_N = \sum_{t=1}^{N} \underline{Y}_{t-1} y_t \quad (3.15)$$

where $\hat{\underline{a}}_N$ $[\hat{a}_N(1), \hat{a}_N(2), \dots, \hat{a}_N(m)]^T$ is the least-squares estimate for $\underline{a}$ and by the subscript N we denote the estimate at time N. Since the matrix in the square brackets in (3.15) can be assumed to be non-singular for sufficiently large N, $\hat{\underline{a}}_N$ is given by

$$\hat{\underset{\sim}{\alpha}}_N = [\sum_{t=1}^{N} \underset{\sim}{Y}_{t-1}\underset{\sim}{Y}_{t-1}^T]^{-1} \sum_{t=1}^{N} \underset{\sim}{Y}_{t-1}y_t . \tag{3.16}$$

However, even though $N \to \infty$, $\hat{\underset{\sim}{\alpha}}_N$ cannot converge in probability to the true value a. Following the similar calculation performed by Sagara & Wada (1977), it follows that

$$\lim_{N \to \infty} \hat{\underset{\sim}{\alpha}}_N - \underset{\sim}{a} = - \underset{\sim}{P}^{-1}\underset{\sim}{a}\,\sigma_n^2 \tag{3.17}$$

with

$$\underset{\sim}{P} = E[\underset{\sim}{Y}_{t-1}\underset{\sim}{Y}_{t-1}^T].$$

In (3.17) $\lim_{N \to \infty} \hat{\underset{\sim}{\alpha}}_N$ implies the stochastic limit of $\hat{\underset{\sim}{\alpha}}_N$. This expression shows that the least-squares estimator is aymptotically biased. Therefore we apply their recursive modified least-squares method to our situation. That is, the estimator $\hat{\underset{\sim}{a}}_N$ is defined by

$$\hat{\underset{\sim}{a}}_N = \hat{\underset{\sim}{\alpha}}_N + N\,\underset{\sim}{P}_N\,\sigma_n^2\,\hat{\underset{\sim}{a}}_{N-1} \tag{3.18}$$

with

$$\underset{\sim}{P}_N = [\sum_{t=1}^{N} \underset{\sim}{Y}_{t-1}\underset{\sim}{Y}_{t-1}^T]^{-1}$$

$$\hat{\underset{\sim}{a}}_N = [\hat{\underset{\sim}{a}}_N(1), \hat{\underset{\sim}{a}}_N(2), \ldots, \hat{\underset{\sim}{a}}_N(m)]^T.$$

It was shown that $\hat{\underset{\sim}{a}}_N$ in (3.18) converges to $\underset{\sim}{a}$ under the assumption that σ_n^2 is known. But in many practical cases, it is necessary to estimate it since σ_n^2 is unknown. By defining the residual at time N as

$$\begin{aligned} \hat{\xi}_t(N) &= y_t - \underset{\sim}{Y}_{t-1}^T\,\hat{\underset{\sim}{\alpha}}_N \\ &= \underset{\sim}{Y}_{t-1}^T(\underset{\sim}{a} - \hat{\underset{\sim}{\alpha}}_N) + \nu_t , \end{aligned} \tag{3.19}$$

the sum of squares of the residuals satisfies the relation

$$\frac{1}{N}\sum_{t=1}^{N}\hat{\xi}_t^2(N) = \frac{1}{N}\sum_{t=1}^{N}\hat{\xi}_t \underset{\sim}{Y}_{t-1}^T(\underset{\sim}{a} - \hat{\underset{\sim}{\alpha}}_N) + \frac{1}{N}\sum_{t=1}^{N}\nu_t^2 \, . \tag{3.20}$$

According to Sagara & Wada (1977), (3.20) converges in probability to

$$\sigma_n^2(\, 1 + \underset{\sim}{a}^T \lim_{N\to\infty} \hat{\underset{\sim}{\alpha}}_N \,) \; + \; \sigma^2$$

as $N \to \infty$. Thus if σ^2 is known, the estimate $(\hat{\sigma_n^2})_N$ of σ_n^2 can be obtained by

$$(\hat{\sigma_n^2})_N = \frac{N^{-1}\sum_{t=1}^{N}\hat{\xi}_t^2(N) - \sigma^2}{1 + \hat{\underset{\sim}{a}}_{N-1}^T\hat{\underset{\sim}{\alpha}}_N} \, . \tag{3.21}$$

Therefore modifying (3.18) to

$$\hat{\underset{\sim}{a}}_N = \hat{\underset{\sim}{\alpha}}_N + N\, \underset{\sim}{P}_N\, (\hat{\sigma_n^2})_N\, \hat{\underset{\sim}{a}}_{N-1} \tag{3.22}$$

and coupling (3.21) with (3.22), consistent estimators of $\underset{\sim}{a}$ and σ_n^2 are obtained. Further the recursive formulas for $\underset{\sim}{P}_N$, $\hat{\underset{\sim}{\alpha}}_N$ and $R_N = \sum_{t=1}^{N}\hat{\xi}_t^2(N)$ are given by

$$\underset{\sim}{P}_{N+1} = \underset{\sim}{P}_N - \underset{\sim}{P}_N\underset{\sim}{Y}_N Q_N^{-1}\underset{\sim}{Y}_N^T\underset{\sim}{P}_N \tag{3.23}$$

$$\hat{\underset{\sim}{\alpha}}_{N+1} \quad \hat{\underset{\sim}{\alpha}}_N + R_N\underset{\sim}{Y}_N(\, y_{N+1} - \underset{\sim}{Y}_N^T\,\hat{\underset{\sim}{\alpha}}_N)Q_N^{-1} \tag{3.24}$$

$$R_{N+1} = R_N + (\, y_{N+1} - \underset{\sim}{Y}_N^T\,\hat{\underset{\sim}{\alpha}}_N)^2\, Q_N^{-1} \tag{3.25}$$

$$Q_N = 1 + \underset{\sim}{Y}_N^T\underset{\sim}{P}_N\underset{\sim}{Y}_N$$

, respectively. Hence, coupling (3.21)-(3.25), the recursive formulas to estimate $\underset{\sim}{a}$ and σ_n^2 have been obtained.

But in many cases, σ^2 is unkown, so that this must be estimated when (3.21) is used. For this reason, we derive the estimator for σ_n^2 not involving σ^2. It follows from (3.19) that

$$\frac{1}{N}\sum_{t=1}^{N}\nu_t\nu_{t-1} = \frac{1}{N}\Big[\sum_{t=1}^{N}\hat{\xi}_t(N)\hat{\xi}_{t-1}(N) - \sum_{t=1}^{N}\hat{\xi}_t(N)\underset{\sim}{Y}_{t-2}^{T}(\underset{\sim}{a}-\hat{\underset{\sim}{\alpha}}_N)$$

$$- \sum_{t=1}^{N}\hat{\xi}_{t-1}(N)\underset{\sim}{Y}_{t-1}^{T}(\underset{\sim}{a}-\hat{\underset{\sim}{\alpha}}_N)$$

$$+ (\underset{\sim}{a}-\hat{\underset{\sim}{\alpha}}_N)^T\sum_{t=1}^{N}\underset{\sim}{Y}_{t-1}\underset{\sim}{Y}_{t-2}^{T}(\underset{\sim}{a}-\hat{\underset{\sim}{\alpha}}_N)\Big]. \quad (3.26)$$

By putting $T_N = \sum_{t=1}^{N}\hat{\xi}_t(N)\hat{\xi}_{t-1}(N)$, the recursive formula for T_N is written as:

$$T_{N+1} = T_N + (y_{N+1} - \underset{\sim}{Y}_N^T\hat{\underset{\sim}{\alpha}}_N)Q_N^{-1}$$

$$\times[y_N - \underset{\sim}{Y}_{N-1}^T\hat{\alpha}_N - \underset{\sim}{Y}_{N-1}^T\underset{\sim}{P}_N\underset{\sim}{Y}_N(y_{N+1} - \underset{\sim}{Y}_N^T\hat{\underset{\sim}{\alpha}}_N)Q_N^{-1}]$$

$$-\sum_{t=1}^{N}[\hat{\xi}_t(N)\underset{\sim}{Y}_{t-2}^T + \hat{\xi}_{t-1}(N)\underset{\sim}{Y}_{t-1}^T]\underset{\sim}{P}_N\underset{\sim}{Y}_N Q_N^{-1}(y_{N+1} - \underset{\sim}{Y}_N^T\hat{\underset{\sim}{\alpha}}_N)$$

$$+ Q_N^{-2}(y_{N+1} - \underset{\sim}{Y}_N^T\hat{\underset{\sim}{\alpha}}_N)^2\underset{\sim}{Y}_N^T\underset{\sim}{P}_N^T\Big(\sum_{t=1}^{N}\underset{\sim}{Y}_{t-1}\underset{\sim}{Y}_{t-2}^T\Big)\underset{\sim}{P}_N\underset{\sim}{Y}_N \quad (3.27)$$

Also denoting $\sum_{t=1}^{N}\underset{\sim}{Y}_{t-1}\underset{\sim}{Y}_{t-2}^T$ and $\sum_{t=1}^{N}[\hat{\xi}_t(N)\underset{\sim}{Y}_{t-2}^T + \hat{\xi}_{t-1}(N)\underset{\sim}{Y}_{t-1}^T]$ in (3.27) by $\underset{\sim}{U}_N$ and $\underset{\sim}{S}_N$, respectively, the recursive forms for $\underset{\sim}{U}_N$ and $\underset{\sim}{S}_N$ are given by

$$\underset{\sim}{U}_{N+1} = \underset{\sim}{U}_N + \underset{\sim}{Y}_N\underset{\sim}{Y}_{N-1} \quad (3.28)$$

$$\underset{\sim}{S}_{N+1} = \underset{\sim}{S}_N + \hat{\xi}_{N+1}(N+1)\underset{\sim}{Y}_{N-1}^T + \hat{\xi}_N(N+1)\underset{\sim}{Y}_N^T$$

$$- \underset{\sim}{Y}_N^T\underset{\sim}{P}_N(\underset{\sim}{U}_N + \underset{\sim}{U}_N^T)(y_{N+1} - \underset{\sim}{Y}_N^T\hat{\underset{\sim}{\alpha}}_N)Q_N^{-1}. \quad (3.29)$$

Hence, (3.26) becomes

$$\frac{1}{N}\sum_{t=1}^{N} \nu_t \nu_{t-1} = \frac{1}{N}[\, T_N - \underset{\sim}{S}_N(\underset{\sim}{a}-\hat{\underset{\sim}{\alpha}}_N) + (\underset{\sim}{a}-\hat{\underset{\sim}{\alpha}}_N)^T \underset{\sim}{U}_N(\underset{\sim}{a}-\hat{\underset{\sim}{\alpha}}_N)\,] \tag{3.30}$$

Moreover, as $N \to \infty$, the left hand side of (3.30) tends to

$$E[\, \nu_t \nu_{t-1} \,] = (-a_1 + a_1 a_2 + \ldots + a_{m-1} a_m)\sigma_n^2. \tag{3.31}$$

Hence we can use the following formula to estimate σ_n^2.

$$(\hat{\sigma}_n^2)_N = \frac{1}{N} \frac{T_N - \underset{\sim}{S}_N(\underset{\sim}{a}-\hat{\underset{\sim}{\alpha}}_N) + (\underset{\sim}{a}-\hat{\underset{\sim}{\alpha}}_N)^T \underset{\sim}{U}_N(\underset{\sim}{a}-\hat{\underset{\sim}{\alpha}}_N)}{-\hat{a}_N(1) + \hat{a}_N(1)\hat{a}_N(2) + \ldots + \hat{a}_N(m-1)\hat{a}_N(m)} . \tag{3.32}$$

So the following recursive formula is considered. That is,

$$\hat{\underset{\sim}{a}}_N = \hat{\underset{\sim}{\alpha}}_N + N\, P_N (\hat{\sigma}_n^2)_{N-1}\, \hat{\underset{\sim}{a}}_{N-1} \tag{3.33}$$

where $(\hat{\sigma}_n^2)_N$ is defined by (3.22). These may be more convenient than using (3.21) and (3.22) because (3.22) and (3.23) do not involve $\hat{\sigma}^2$ and are closed with respect to $\hat{\underset{\sim}{a}}_N$ and $(\hat{\sigma}_n^2)_N$. Furthermore from (3.21), the estimate $(\hat{\sigma}^2)_N$ of σ^2 is given by

$$(\hat{\sigma}^2)_N = \frac{1}{N} R_N - (\hat{\sigma}_n^2)_N (1 + \hat{\underset{\sim}{a}}_{N-1}^T \hat{\underset{\sim}{\alpha}}_N). \tag{3.34}$$

We have presented the MLS estimator.

Before deriving the error covariance for the MLS estimate, we describe shortly about the estimation error for the least-squares estimator $\hat{\underset{\sim}{\alpha}}_N$. As $N \to \infty$, (3.15) becomes $E[\underset{\sim}{Y}_{t-1}\underset{\sim}{Y}_{t-1}^T]\, \underset{\sim}{\alpha} = E[\underset{\sim}{Y}_{t-1} y_t]$ where $\underset{\sim}{\alpha} = \lim_{N\to\infty} \hat{\underset{\sim}{\alpha}}_N$. By the same argument in Section 2.2, the estimation error $\Delta\underset{\sim}{\alpha} = \hat{\underset{\sim}{\alpha}}_N - \underset{\sim}{\alpha}$ is asymptotically expressed

$$(\underset{\sim}{P}\Delta\underset{\sim}{\alpha})_k \sim \int_{-\pi}^{\pi} I_N(s)\, [\, 1 - \sum_{i=1}^{m} \alpha_i\, e^{-jis}\,]\, e^{jks}\, ds. \tag{3.35}$$

Now we derive the error covariance for the MLS estimate by the same argument as above. Instead of treating (3.33) directly, we perform the error analysis for the following estimator;

$$\hat{\underline{a}}_N = \hat{\underline{\alpha}}_N + N\,\underline{P}_N\,(\hat{\sigma}_n^2)_N\,\hat{\underline{a}}_N$$

or

$$\hat{\underline{a}}_N = [\,\underline{I} - N\,\underline{P}_N\,(\hat{\sigma}_n^2)\,]^{-1}\,\hat{\underline{\alpha}}_N \qquad (3.36)$$

In asymptotical sense, the result for (3.36) may be identical with the one for (3.33). Since the calculation for obtaining a is rather complicated, we shall describe it in Appendix 3.1. The result is

$$\underline{A}\,\Delta\underline{a} - \int_{-\pi}^{\pi} I_N(s)\,\underline{g}(s)\,ds$$

where $\underline{A} = \underline{I} - \underline{P}^{-1}\sigma_n^2 - \underline{P}^{-1}\underline{a}\,\partial\sigma_n^2/\partial\underline{a}$, and $\underline{g}(s)$ is defined by (A-3.6). We can easily show that

$$\int_{-\pi}^{\pi} f_y(s)\,g(s)\,ds = \underline{0}$$

Thus, we have

$$\underline{A}\,E[\Delta\underline{a}\Delta\underline{a}^T]\,\underline{A}^T \simeq \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} g(s)g^T(t)\mathrm{Cov}[I_N(s),I_N(t)]ds\cdot dt. \qquad (3.37)$$

Substitution of $\mathrm{Cov}[I_N(s),I_N(t)] \simeq 2\pi N^{-1}f_y(s)f_y(t)[\delta(s+t)+\delta(s-t)]$ into (3.37) gives

$$\underline{A}\,E[\Delta\underline{a}\Delta\underline{a}^T]\,\underline{A}^T \simeq 2\pi N^{-1}\int_{-\pi}^{\pi} f^2(s)[\,\underline{g}(s)\underline{g}^T(s) + \underline{g}(s)\underline{g}^T(-s)]ds$$

$$= N^{-1}\underline{\Gamma} \qquad (3.38)$$

3.4 Comparison of Two Estimators and Simulations

In this section, we first compare the performances of the above two estimators by calculating (3.13) and (3.38) for the first and second order AR processes. We shall show that for the first order AR process, the MLS estimate is identical with the Y-W estimate i.e., both error covariances coincide each other. These evaluations are simple because all the equations reduce to scalar forms. From (3.36) and (3.32), it follows that

$$\begin{aligned} P_N U_N \hat{a}_N^2 &+ (1 - S_N P_N - 2 \hat{\alpha}_N U_N P_N) \hat{a}_N \\ &+ (-\hat{\alpha}_N + P_N T_N + S_N \hat{\alpha}_N P_N + \hat{\alpha}_N^2 U_N P_N) = 0 \end{aligned} \tag{3.39}$$

where

$$T_N = N(\hat{r}_1 - \hat{r}_2 \hat{\alpha}_N - \hat{r}_0 \hat{\alpha}_N + \hat{r}_1 \hat{\alpha}_N^2), \; U_N = N\hat{r}_1.$$

$$S_N = N(\hat{r}_2 - \hat{r}_1 \hat{\alpha}_N + \hat{r}_0 - \hat{r}_1 \hat{\alpha}_N), \; P_N = 1/(N\hat{r}_0)$$

Substituting T_N, S_N, U_N and P_N into (3.39), coefficients of $\hat{a}_N^2$, $\hat{a}_N$ and the constant term are $\hat{r}_1/\hat{r}_0$, $-\hat{r}_2/\hat{r}_0$, and zero, respectively. Therefore, the solution a_N of (3.39) is $\hat{r}_2/\hat{r}_1$, which is also obtained from (3.6). To check the validity of the formula (3.38), we apply it for this case. From (3.38) we have

$$\begin{aligned} A^2 E[(\Delta a)^2] = &\frac{1}{2\pi N} \int_{-\pi}^{\pi} [\frac{\sigma^2}{B(s)B(-s)} + \sigma_n^2]^2 g^2(s) \, ds \\ &+ \frac{1}{2\pi N} \int_{-\pi}^{\pi} [\frac{\sigma^2}{B(s)B(-s)} + \sigma_n^2]^2 g(s)g(-s) ds. \end{aligned} \tag{3.40}$$

From (A-3.6), $g(s)$ becomes

$$g(s) = \frac{1}{r_0} \{ \frac{a\sigma_n^2}{r_0} + [1 + e^{j2s}]a - a^2 e^{js} - \alpha \}.$$

From (3.15) we obtain $\alpha = r_1/r_0 = a - a\sigma_n^2/r_0$. Thus $g(s)$ turns into $g(s) = a \exp(j2s)B(s)/r_0$. Therefore the first term of (3.40) is

$$\frac{1}{2\pi Nj}\frac{a^2}{r_0^2}\oint_{|z|=1}\left[\frac{\sigma^4 z}{B_0^2(z)} + \frac{2\sigma^2\sigma_n^2 a^2}{B_0(z)} + \sigma_n^4 B_1(z)z\right] z^2\,dz$$

where $B_0(z) = 1 - az$ and $B_1(z)z = z - a$. Since by the stationary assumption, the root of $B_0(z) = 0$ lies outside the unit circle, the integrand of the above complex integral is regular within the unit circle, thereby the integral is zero from Cauchy's theorem. The second term is given by

$$\frac{1}{Nr_0^2}\left[\frac{a^2\sigma^4}{1-a^2} + 2\,\sigma^2\sigma_n^2\,a^2 + \sigma_n^4 a^2(1+a^2)\right]. \quad (3.41)$$

On the other hand, A is given by

$$A = 1 - \frac{\sigma_n^2}{r_0} - \frac{a}{r_0}\frac{\partial\sigma_n^2}{\partial a}.$$

But from (A-3.4), $\partial\sigma_n^2/\partial a = (r_2 + r_0 - 2r_1 a - \sigma_n^2)/a = \sigma^2/a$, so that A becomes $a^2\sigma^2/r_0(1-a^2)$. Thus, we finally obtain the following equation.

$$N\,E[(\Delta a)^2] \approx \frac{1-a^2}{a^2}\left[1 + 2(1-a^2)\frac{\sigma_n^2}{\sigma^2} + (1-a^4)\frac{\sigma_n^4}{\sigma^4}\right] \quad (3.42)$$

It is seen that (3.42) is identical with the result (3.14).

Next, we consider the second order AR process. From (3.13), the error covariance for the Y-W estimator is obtained as follows:

$$N\,E[(\Delta a_1)^2] \approx \frac{1}{(r_2^2 - r_1r_3)^2}[(r_2^2 + r_1^2)h_1 - 2r_1r_2h_2] \quad (3.43)$$

$$N\,E[(\Delta a_2)^2] \approx \frac{1}{(r_2^2 - r_1r_3)^2}[(r_3^2 + r_2^2)h_1 - 2r_2r_3h_2] \quad (3.44)$$

where

$$h_1 = e_{1,1}\,\sigma_n^4 + 2\sigma_n^2\sigma^2 + \sigma^2 r_0'$$

$$h_2 = e_{1,2}\,\sigma_n^4 + \sigma^2 r_1'$$

$$e_{1,1} = 1 + a_1^2 + a_2^2, \quad e_{1,2} = -a_1 + a_1 a_2$$

Since the detailed calculation of (3.38) is straightforward but lengthy, we omit it and only describe the main current of the calculation. In the first step, we substitute the relation (3.17), that is, $\underset{\sim}{P}^{-1}\underset{\sim}{a} = (\underset{\sim}{a} - \underset{\sim}{\alpha})/\sigma_n^2$ into $\underset{\sim}{g}(s)$ to make this calculation easy. Then rearranging the terms with respect to the powers of $\exp(js)$, we obtain

$$\underset{\sim}{g}(s) = \begin{bmatrix} G_0 + G_1 e^{js} + G_2 e^{j2s} + G_3 e^{j3s} + G_{-1} e^{-js} \\ H_0 + H_1 e^{js} + H_2 e^{j2s} + H_3 e^{j3s} + H_{-1} e^{-js} \end{bmatrix}$$

where G_i's and H_i's are some constants. Further to calculate the integral in (3.38), the following relation is used;

$$\int_{-\pi}^{\pi} f_y^2(s)\, e^{jis}\, ds = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} r_k\, r_{k-i} \qquad (3.45)$$

Consequently, the elements of the matrix $\underset{\sim}{\Gamma}$ are as follows:

$$\underset{\sim}{\Gamma}_{1,1} = \sum_{i=0}^{6} \gamma_i \sum_{k=-\infty}^{\infty} r_k\, r_{k-i}$$

$$\underset{\sim}{\Gamma}_{1,2} = \underset{\sim}{\Gamma}_{2,1} = \sum_{i=0}^{6} \eta_i \sum_{k=-\infty}^{\infty} r_k\, r_{k-i}$$

$$\underset{\sim}{\Gamma}_{2,2} = \sum_{i=0}^{6} \zeta_i \sum_{k=-\infty}^{\infty} r_k\, r_{k-i}$$

where γ_i's, η_i's and ζ_i's are complicated function of G_i's and H_i's and in actual calculations the infinite sums are truncated appropriately because the autocovariance function is a mixture of damped exponentials and damped sine waves.

However, the above results are rather complicated so that we calculate them numerically for some examples. Table 3.1 and 3.2 show the simulation results of the error variances calculated from 50 sample paths each of 2000 length along with the theoretical values (3.43), (3.44), (3.38) and the corresponding Cramér-Rao bound, where (3.45) is truncated at $k = \pm 80$. The SNR is defined by σ_x^2/σ_n^2, where σ_x^2 is the average power of the signal $\{x_t\}$ and is given by $(1-a_2)\sigma^2/(1+a_2)\{(1-a_2)^2-a_1\}$. In this case, the information matrix is obtained by Pagano (1974) as

$$\frac{1}{4\pi}\int_{-\pi}^{\pi} \underset{\sim}{\theta}(s)\, \underset{\sim}{\theta}^T(s)\, ds$$

where

$$\underset{\sim}{\theta}^T(s) = \left(\frac{\partial}{\partial a_1}, \frac{\partial}{\partial a_2}, \frac{\partial}{\partial \sigma_n^2}, \frac{\partial}{\partial \sigma^2} \right) \ln f_y(s).$$

SNR		3.9		15.6	
		$NE[(\Delta a_1)^2]$	$NE[(\Delta a_2)^2]$	$NE[(\Delta a_1)^2]$	$NE[(\Delta a_2)^2]$
Y-W	Simulation	296.73	265.88	70.31	63.26
	(3.43),(3.44)	163.60	146.69	47.16	42.64
MLS	Simulation	6.33	5.55	1.81	1.58
	(3.38)	5.07	4.33	1.66	1.44
Cramér-Rao bound		4.93	4.20	1.64	1.43

Table 3.1 Comparison of the error variances between simulation results and theoretical values for $a_1 = 0.5$ and $a_2 = 0.4$.

SNR		1.8		7.1	
		$NE[(\Delta a_1)^2]$	$NE[(\Delta a_2)^2]$	$NE[(\Delta a_1)^2]$	$NE[(\Delta a_2)^2]$
Y-W	Simulation	10.44	7.28	6.63	2.24
	(3.43), (3.44)	13.10	6.91	6.64	2.25
MLS	Simulation	7.48	4.48	2.59	2.12
	(3.38)	6.65	4.37	2.52	1.99
Cramer-Rao bound		4.80	3.40	2.44	1.93

Table 3.2 Comparison of the error variances between simulation results and theoretical values for $a_1 = 0.75$ and $a_2 = -0.5$.

From these tables, it may be seen that simulation results and theoretical values have similar trends. Table 3.1 shows an example that the MLS estimate is much better than the Y-W estimate, and Table 3.2 shows an example that the MLS estimate is slightly better.

For the BS estimate, it is difficult to obtain the corresponding theoretical values, so that only the simulation results are presented. For $N\ E[(\Delta a_1)^2]$, $N\ E[(\Delta a_2)^2]$, experimental values are 221.33, 263.13 with SNR = 3.9 and 182.69, 209.73 with SNR = 15.6, respectively. Thus in these cases, the performances of the BS estimate is nearly equal to that of the Y-W estimate.

Next, in Fig. 3.1 and 3.2, comparison between the theoretical error variances of both methods are made for all the values of a_1

a_2 guaranteeing the stationality of the process with SNR = 2 fixed. It is well known that this area is a triangular region as is described in Box & Jenkins (1970). As can be seen from (3.43), (3.44) and (3.38), the error variances are symmetric about $a_1 = 0$, so that only the half of the triangular is considered. In these figures, the symbol ⊖ implies that the error variances of the MLS estimator are much less than the corresponding values of the Y-W estimator, that is, the former is much better. O implies that the former is slightly better and implies that the former is slightly worse. In practical sense, both estimators have same performance in the region marked by O and ×. Thus the MLS estimator has always much better or equal performance than the Y-W estimator. It can be seen that O and × are mainly concentrated on the area below the parabolic boundary $a_1^2 + 4a_2 < 0$ where the autocovariance function is a damped sine wave. It is difficult to understand the exact reason for this phenomenon. But we might have a rather qualitative reasoning. In this second order case, the highest lag of the autocovariances used is 3 for the MLS estimator and 4 for the Y-W estimator, respectively. In general, the estimate of a higher lag autocovariance has less efficiency than that of a lower one. However, when the shape of an autocovariance function is similar to a purely cosine wave, this is not so. Hence , below the parabolic boundary, the use of third and fourth lag estimates does not give a clear difference, whereas, in the overdamped region, this causes a great difference. So far, we have stated relative comprisons of the two estimators. Next, absolute

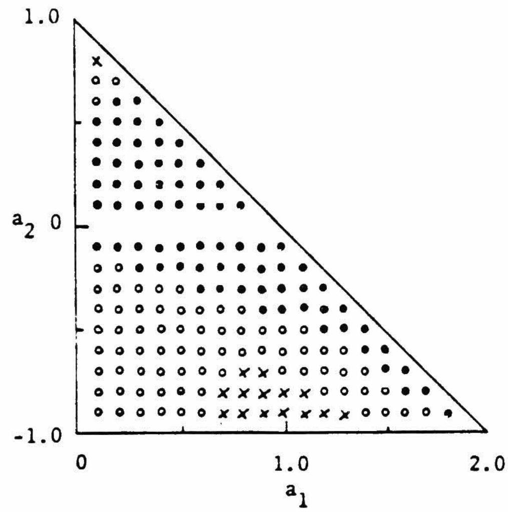

Fig. 3.1 Comparison between the theoretical error variances of MLS and Y-W methods for a_1.

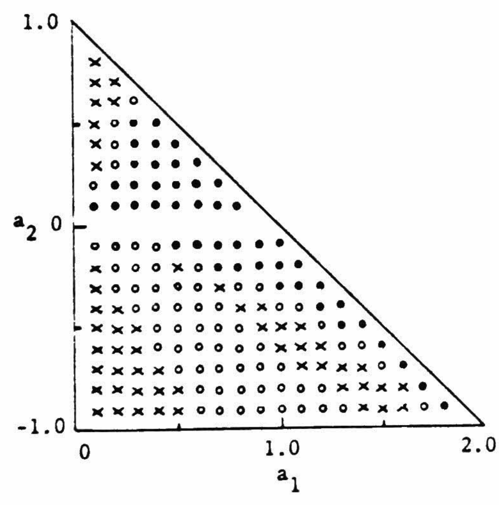

Fig. 3.2 Comparison between the theoretical error variances of MLS and Y-W methods for a_2 .

comparisons are briefly mentioned. Let the relative efficiency be defined by the ratio of an asymptotic variance of some method to the corresponding Cramér-Rao bound. For the examples of Table 3.1 and 3.2, the efficiencies of the MLS estimator for a_1 amount to 97.2 % with SNR = 3.9, 72.3 % with SNR = 1.8, respectively. However, this high or moderate performance does not prevail over all regions, for at $a_1 = 1.8$, $a_2 = -0.9$, the efficiency for a_1 with SNR = 2 is seriously degraded to 0.3 %. (The corresponding efficiency for the Y-W estimate is 0.09 %.) The reason of this phenomenon will be discussed in Chapter 5.

Finally, Figs.3.3-3.6 show sample runs of the MLS and the Y-W recursive estimators with SNR = 10 to see their long range behaviours. In the early stage of estimation, the fluctuations of the Y-W estimator are relatively large as compared with the MLS estimate. This can be also inferred from the theoretical error variances.

3.5 A Remark and Conclusion

We note that the MLS algorithm can be applied to the case where the observations are occasionally interrupted, or more precisely , are expressed as

$$y_t = d_t x_t + n_t \tag{3.46}$$

where $\{d_t\}$ is a Bernoulli sequence with $Pr(d_t=1)$ p and $Pr(d_t=0)$ =1-p. We explain the reason as follows. From (3.46) we have

$$r_k = \begin{cases} p\, r'_0 + \sigma_n^2 & \text{for } k = 0 \\ p^2 r'_k & \text{for } k \neq 0. \end{cases} \tag{3.47}$$

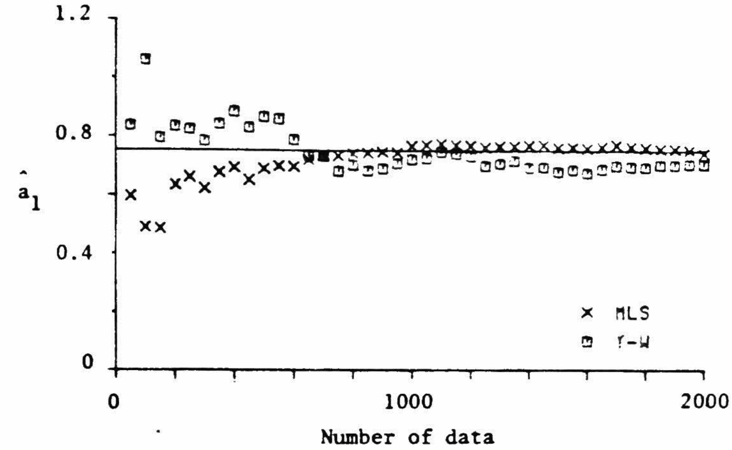

Fig. 3.3 Sample runs of recursive MLS and Y-W estimates for a_1 of the process $x_t = 0.75x_{t-1} + 0.5x_{t-2} + u_t$.

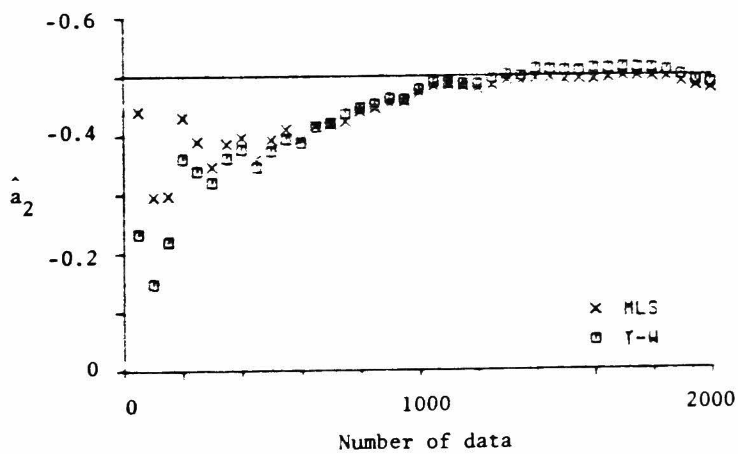

Fig. 3.4 Sample runs of recursive MLS and Y-W estimates for a_2 of the process $x_t = 0.75x_{t-1} + 0.5x_{t-2} + u_t$.

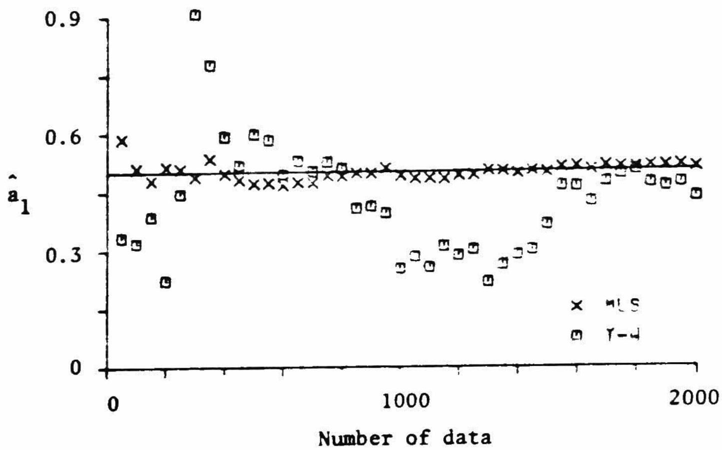

Fig. 3.5 Sample runs of recursive MLS and Y-W estimates for a_1 of the process $x_t = 0.5x_{t-1} + 0.4x_{t-2} + u_t$.

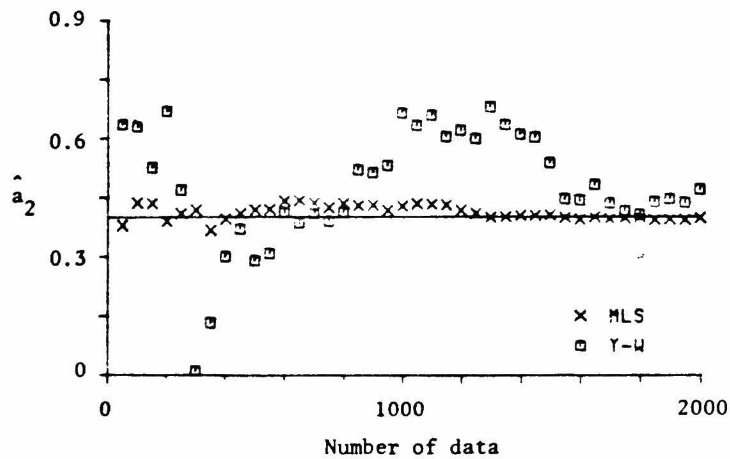

Fig. 3.6 Sample runs of recursive MLS and Y-W estimates for a_2 of the process $x_t = 0.5x_{t-1} + 0.4x_{t-2} + u_t$.

But noting that $r_0 = p^2 r'_0 + (p-p^2) r'_0 + \sigma_n^2 = p^2 r'_0 + \sigma_{n'}^2$, the following system ;

$$x'_t - a_1 x'_{t-1} - \cdots - a_m x'_{t-m} = u'_t$$

$$y_t = x'_t + n'_t \qquad (3.48)$$

$$E[u'_t u'_s] = p^2 \sigma^2 \delta_{t,s}, \quad E[u'_t n'_s] = 0, \ E[n'_t n'_s] = \sigma_{n'}^2 \delta_{t,s}$$

has the same sequence of autocovariances with $\{r_k\}$ in (3.47). Thus the MLS algorithm based on (3.46) yields a consistent estimate for $\underset{\sim}{a}^T = (a_1, \ldots, a_m)$ even if p is *unknown*. But if so, we cannot estimate one of σ^2, σ_n^2, and p in a simple manner. We will extensively study statistical problems associated with missing observations in the next chapter.

In conclusion, in this chapter, we have discussed three methods , that is, the Y-W, the MLS, and the BS methods, for the problem of estimating AR process parameters based on the data corrupted by unkown white noise. It may be concluded that on the whole, the MLS estimator is best among the three. Moreover, the formulas of the asymptotic theoretical error covariance matrices have been justified by simulations. The MLS method has potential applications in speech and image processing where restoration of noisy speech or image is executed by Kalman filtering after estimating the parameters.

Appendix 3.1

In this appendix, we derive the asymptotic expression for

$\Delta\underset{\sim}{a}$ in (3.36). Substituting $(\hat{\sigma}_n^2)_N = \sigma_n^2 + \Delta\sigma_n^2$, $\hat{\underset{\sim}{\alpha}}_N = \underset{\sim}{\alpha} + \Delta\underset{\sim}{\alpha}$ and $N\underset{\sim}{P}_N = \underset{\sim}{P}^{-1} + \Delta\underset{\sim}{P}^{-1}$ into (3.36) and neglecting the second order terms concerning the errors, we have

$$\hat{\underset{\sim}{a}}_N = [\ \underset{\sim}{I} - (\underset{\sim}{P}^{-1} + \Delta\underset{\sim}{P}^{-1})(\sigma_n^2 + \Delta\sigma_n^2)\ ]^{-1}(\underset{\sim}{\alpha} + \Delta\underset{\sim}{\alpha})$$

$$= [\ \underset{\sim}{I} - \underset{\sim}{P}^{-1}\sigma_n^2 - \Delta\underset{\sim}{P}^{-1}\sigma_n^2 - \underset{\sim}{P}^{-1}\Delta\sigma_n^2\ ]^{-1}(\underset{\sim}{\alpha} + \Delta\underset{\sim}{\alpha}).$$

Then using the relation $(\underset{\sim}{M} - \Delta\underset{\sim}{M})^{-1} - \underset{\sim}{M}^{-1} + \underset{\sim}{M}^{-1}(\Delta\underset{\sim}{M})\underset{\sim}{M}^{-1}$, the above equation turns into

$$\hat{\underset{\sim}{a}}_N = [(\underset{\sim}{I} - \underset{\sim}{P}^{-1}\sigma_n^2)^{-1} + (\underset{\sim}{I} - \underset{\sim}{P}^{-1}\sigma_n^2)^{-1}(\Delta\underset{\sim}{P}^{-1}\sigma_n^2 + \underset{\sim}{P}^{-1}\Delta\sigma_n^2)$$

$$\times(\underset{\sim}{I} - \underset{\sim}{P}^{-1}\sigma_n^2)^{-1}](\underset{\sim}{\alpha} + \Delta\underset{\sim}{\alpha}\ )$$

$$= \underset{\sim}{a} + (\underset{\sim}{I} - \underset{\sim}{P}^{-1}\sigma_n^2)^{-1}(\ \underset{\sim}{P}^{-1}\sigma_n^2 + \underset{\sim}{P}^{-1}\Delta\sigma_n^2)\underset{\sim}{a} + (\underset{\sim}{I} - \underset{\sim}{P}^{-1}\sigma_n^2)^{-1}\Delta\underset{\sim}{\alpha}.$$

Also substituting the following relation

$$\Delta\underset{\sim}{P}^{-1} = N\underset{\sim}{P}_N - \underset{\sim}{P}^{-1} = \hat{\underset{\sim}{P}}^{-1} - \underset{\sim}{P}^{-1} = (\hat{\underset{\sim}{P}} - \underset{\sim}{P} + \underset{\sim}{P})^{-1} - \underset{\sim}{P}^{-1}$$

$$= -\underset{\sim}{P}^{-1}\hat{\underset{\sim}{P}}\,\underset{\sim}{P}^{-1} + \underset{\sim}{P}^{-1}.$$

, we have

$$(\underset{\sim}{I} - \underset{\sim}{P}^{-1}\sigma_n^2)\Delta\underset{\sim}{a} = (\sigma_n^2\underset{\sim}{I} - \underset{\sim}{P}^{-1}\hat{\underset{\sim}{P}}\,\sigma_n^2 + \Delta\sigma_n^2\,\underset{\sim}{I})\underset{\sim}{P}^{-1}\underset{\sim}{a} + \Delta\underset{\sim}{\alpha} \quad \text{(A-3.1)}$$

where $\Delta\underset{\sim}{a} = \hat{\underset{\sim}{a}}_N - \underset{\sim}{a}$. Since (3.32) is rewritten as

$$(\hat{\sigma}_n^2)_N = \frac{\hat{T} - \hat{\underset{\sim}{S}}(\underset{\sim}{a}_N - \underset{\sim}{\alpha}_N) + (\underset{\sim}{a}_N - \underset{\sim}{\alpha}_N)^T\hat{\underset{\sim}{U}}(\underset{\sim}{a}_N - \underset{\sim}{\alpha}_N)}{-a_N(1) + a_N(1)a_N(2) + \dots + a_N(m-1)a_N(m)} \quad \text{(A-3.2)}$$

with $\hat{T} = N^{-1}T_N$, $\hat{\underset{\sim}{S}} = N^{-1}\underset{\sim}{S}_N$ and $\hat{\underset{\sim}{U}} = N^{-1}\underset{\sim}{U}_N$, the estimation error $\Delta\sigma_n^2$ of σ_n^2 is asymptotically given by

$$\Delta\sigma_n^2 \approx \frac{\partial\sigma_n^2}{\partial \underset{\sim}{a}}\Delta\underset{\sim}{a} + \frac{\partial\sigma_n^2}{\partial \underset{\sim}{\alpha}}\Delta\underset{\sim}{\alpha} + \frac{\partial\sigma_n^2}{\partial T}\Delta T + \Delta\underset{\sim}{S}\left(\frac{\partial\sigma_n^2}{\partial \underset{\sim}{S}}\right)^T + \mathrm{tr}\left[\left(\frac{\partial\sigma_n^2}{\partial \underset{\sim}{U}}\right)^T \Delta\underset{\sim}{U}\right] \qquad \text{(A-3.3)}$$

where we define

$$\Delta T = \hat{T} - T$$

$$\Delta\underset{\sim}{S} = \hat{\underset{\sim}{S}} - \underset{\sim}{S}$$

$$\Delta\underset{\sim}{U} = \hat{\underset{\sim}{U}} - \underset{\sim}{U}$$

$$\lim_{N\to\infty} \hat{T} = T = E[\ \hat{\xi}_t(N)\hat{\xi}_{t-1}(N)\ ]$$

$$\lim_{N\to\infty} \hat{\underset{\sim}{S}} = \underset{\sim}{S} = E[\ \hat{\xi}_t(N)\underset{\sim}{Y}_{t-2}^T + \hat{\xi}_{t-1}(N)\underset{\sim}{Y}_{t-1}^T\ ]$$

and

$$\lim_{N\to\infty} \hat{\underset{\sim}{U}} = \underset{\sim}{U} = E[\ \underset{\sim}{Y}_{t-1}\underset{\sim}{Y}_{t-2}^T\ ].$$

In (A-3.3), the term like $\Delta\underset{\sim}{S}(\partial\sigma_n^2/\partial\underset{\sim}{S})^T$ means the partial variation due to the slight change of $\underset{\sim}{S}$. From (A-3.2), the term in (A-3.3) can be expressed as follows :

$$\frac{\partial\sigma_n^2}{\partial \underset{\sim}{a}}\Delta\underset{\sim}{a} = \frac{1}{c}\left[\ -\underset{\sim}{S} + (\underset{\sim}{a} - \underset{\sim}{\alpha})^T(\underset{\sim}{U}^T + \underset{\sim}{U}) - \sigma_n^2\underset{\sim}{\Lambda}\ \right]\Delta\underset{\sim}{a}$$

$$\frac{\partial\sigma_n^2}{\partial \underset{\sim}{\alpha}}\Delta\underset{\sim}{\alpha} = \frac{1}{c}\left[\ \underset{\sim}{S}\Delta\underset{\sim}{\alpha} + (\underset{\sim}{\alpha} - \underset{\sim}{a})^T(\underset{\sim}{U}^T + \underset{\sim}{U})\Delta\underset{\sim}{\alpha}\ \right]$$

$$\mathrm{tr}\left[\left(\frac{\partial\sigma_n^2}{\partial \underset{\sim}{U}}\right)^T \Delta\underset{\sim}{U}\right] = \frac{1}{c}(\underset{\sim}{a} - \underset{\sim}{\alpha})^T\,\Delta\underset{\sim}{U}(\underset{\sim}{a} - \underset{\sim}{\alpha})$$

$$\frac{\partial\sigma_n^2}{\partial T}\Delta T = \frac{1}{c}\cdot\Delta T, \qquad \Delta\underset{\sim}{S}\left(\frac{\partial\sigma_n^2}{\partial \underset{\sim}{S}}\right)^T = -\frac{1}{c}\cdot\Delta\underset{\sim}{S}(\underset{\sim}{a} - \underset{\sim}{\alpha})$$

where

$$c = -a_1 + a_1a_2 + \dots + a_{m-1}a_m$$

$$\underset{\sim}{\Lambda} = [\ -1+a_2,\ a_1+a_3,\ \dots,\ a_{m-2}+a_m,\ a_{m-1}\ ].$$

By substituting (A-3.3) into (A-3.1) and rearranging the terms, it follows that

$$(\underset{\sim}{I} - \underset{\sim}{P}^{-1}\sigma_n^2 - \underset{\sim}{P}^{-1}\underset{\sim}{a}\frac{\partial\sigma_n^2}{\partial \underset{\sim}{a}}) \Delta\underset{\sim}{a}$$

$$= \{ \sigma_n^2 \underset{\sim}{I} - \underset{\sim}{P}^{-1}\hat{\underset{\sim}{P}}\sigma_n^2 + \frac{\partial\sigma_n^2}{\partial \underset{\sim}{\alpha}}\Delta\underset{\sim}{\alpha} + \frac{\partial\sigma_n^2}{\partial T}\Delta T + \Delta\underset{\sim}{S}(\frac{\partial\sigma_n^2}{\partial \underset{\sim}{S}})^T$$

$$+ \mathrm{tr}[(\frac{\partial\sigma_n^2}{\partial \underset{\sim}{U}})^T\Delta\underset{\sim}{U}]\} \underset{\sim}{P}^{-1}\underset{\sim}{a} + \Delta\underset{\sim}{\alpha}. \quad \text{(A-3.5)}$$

Next, substituting (A-3.4) into the terms in the brace in the right hand side of (A-3.5) and using the explicit equation for ΔT, $\Delta\underset{\sim}{S}$, and $\Delta\underset{\sim}{U}$, it follows that

$$-\underset{\sim}{P}^{-1}\hat{\underset{\sim}{P}}\sigma_n^2 + \frac{1}{c}\frac{1}{N}\{\sum_{t=1}^{N} y_t y_{t-1} - \sum_{t=1}^{N}(y_t\underset{\sim}{Y}_{t-2}^T + y_{t-1}\underset{\sim}{Y}_{t-1}^T)\underset{\sim}{a}$$

$$+ \underset{\sim}{a}^T \sum_{t=1}^{N} \underset{\sim}{Y}_{t-1}\underset{\sim}{Y}_{t-2}^T \underset{\sim}{a}\}$$

where the process of the above derivation is omitted because it is rather lengthy. Rewriting (A-3.5) in terms of the periodogram , we finally obtain

$$(\underset{\sim}{I} - \underset{\sim}{P}^{-1}\sigma_n^2 - \underset{\sim}{P}^{-1}\underset{\sim}{a}\frac{\partial\sigma_n^2}{\partial \underset{\sim}{a}}) \Delta\underset{\sim}{a} = \int_{-\pi}^{\pi} I_N(s)$$

$$\times \left\{ e^{js} - c\underset{\sim}{P}^{-1}\begin{bmatrix} 1 & \cdots & e^{j(m-1)s} \\ e^{js} & & \vdots \\ \vdots & & \vdots \\ e^{j(m-1)s} & \cdots & 1 \end{bmatrix} \right.$$

$$- \begin{bmatrix} e^{j2s} & + & 1 \\ e^{j3s} & + & e^{js} \\ & \vdots & \\ e^{j(m+1)s} & + & e^{j(m-1)s} \end{bmatrix}^T \underset{\sim}{a}$$

$$+ \quad \underset{\sim}{a}^T \begin{bmatrix} e^{js} & e^{j2s} & \cdots & e^{jms} \\ 1 & \cdots & & \vdots \\ e^{js} & & & \vdots \\ \vdots & & & \\ e^{j(m-2)s} & & \cdots & e^{js} \end{bmatrix} \underset{\sim}{a} \Bigg\} \frac{1}{c} \underset{\sim}{P}^{-1} \underset{\sim}{a} \, ds$$

$$+ \int_{-\pi}^{\pi} I_N(s) \, P^{-1} [\, 1 \quad \sum_{i=1}^{m} \alpha_i \, e^{-jis} \,] \cdot \begin{bmatrix} e^{js} \\ e^{j2s} \\ \vdots \\ e^{jms} \end{bmatrix} ds$$

$$\triangleq \int_{-\pi}^{\pi} I_N(s) \, \underset{\sim}{g}(s) \, ds. \qquad \text{(A-3.6)}$$

Chapter 4

Fitting Autoregression with Missing Observations

4.1 Introduction

The problem of time series with missing observations was first treated by Jones (1962) where the missing observations were assumed to occur periodically and their effect on spectral analysis was investigated. After this work, several papers concerning this aspect, for example, Parzen (1963), Scheinok (1965) , Bloomfield (1970), Neave (1970), Jones (1970), and Alekseev & Savitsky (1973) have been published, but all these papers are concerned with the analysis in the frequency domain, or in other words, spectral analysis based on variously modified Blackman-Tukey procedures.

As far as the author is aware, no results have ever been obtained concerning the effects of missing observations on the estimates of parameteric models.

In this chapter, we derive the expressions for the error covariance matrices of the estimate of AR process parameters based on the data with randomly and regularly missed observations by using the periodogram technique. Although we are apt to think that missing observations always have negative effects on estimating the parameters, in some cases, we can positively utilize the concept of missing observations to decrease the variances if the number of observations is fixed but time instances at which the observations are made can be changed.

4.2 Randomly Missed Observations

Scheinok (1965) considered the case where missing observations occur stochastically and the missing instants form a Bernoulli sequence. In this section, we set the same situation, namely,

$$d_t = \begin{cases} 1 & \text{if } x_t \text{ is read} \\ 0 & \text{if } x_t \text{ is not read} \end{cases}$$

where the time series $\{x_t\}$ is a Gaussian m-th order AR process (2.1) and the d_t's are independent not only with each other but also with the time series $\{x_t\}$ with $p = \Pr(d_t = 1)$ known *a priori*. We also denote $r_k = E[x_t x_{t+k}]$. With the above assumptions, it is obvious that the estimators for the r_k's

$$\begin{aligned} \hat{r}'_0 &= \frac{1}{N\,p} \sum_{t=1}^{N} (d_t x_t)^2 \\ \hat{r}'_k &= \frac{1}{Np^2} \sum_{t=1}^{N-|k|} d_t x_t d_{t+|k|} x_{t+|k|} \quad k \neq 0 \end{aligned} \tag{4.1}$$

are consistent as $N \to \infty$. Hence, substitution of (4.1) into (2.5) and (2.14) gives the consistent estimators $\hat{\underset{\sim}{a}}'$, $\hat{\sigma}^{2\prime}$ for $\underset{\sim}{a}$ and σ^2, respectively. Corresponding to (2.9), the modified periodogram is now defined as in Scheinok (1965) by

$$I'_N(s) = (2\pi N)^{-1} \{ \sum_{i=1}^{N} \frac{d_i^2 x_i^2}{p} + \sum_{k=1}^{N} \sum_{i=1;i\neq k}^{N} \frac{d_k d_i}{p^2} x_k x_i e^{-j(i-k)s} \}. \tag{4.2}$$

Then, corresponding to (2.9), the following relation between $\hat{r}'_k$ and $I'_N(s)$ holds:

$$\hat{r}'_k = \int_{-\pi}^{\pi} I'_N(s)\, e^{jks}\, ds. \qquad (4.3)$$

Also, by the assumption concerning $\{d_t\}$ and (4.2), it is obvious

$$E[\, I'_N(s)\,] = E[\, I_N(s)\,].$$

where $I_N(s)$ is the usual periodogram without missing observations. Hence, all the arguments in Chapter 2 are entirely valid by interchanging $I_N(s)$ with $I'_N(s)$. Thus it suffices to know $\mathrm{Cov}[I'_N(s),I'_N(t)]$ but its derivation is just the central theme of Scheinok (1965). The result is

$$\begin{aligned}
\mathrm{Cov}[I'_N(s),I'_N(t)] \;\doteq\;& 3(4\pi^2N)^{-1}\underline{(p^{-1}-1)}v + (\pi N)^{-1}w \\
& + (\pi N)^{-1}(3p^{-1}-1)[f(s)+f(s)-\pi^{-1}v] + 2N^{-1}\{f^2(s) \\
& + f^2(t) - \pi^{-1}\int_{-\pi}^{\pi} f(x)[f(x)+f(s)+f(t)]dx + \pi^{-2}\alpha^2\} \\
& + [f(s)f(t)]N^{-2}[F_N(s+t) + F_N(s-t)] - 2N^{-1}[f(s)+f(t)]^2 \\
& - (2\pi N^2)^{-1}[F_N(s+t) + F_N(s-t)][f(s)+f(t)]v + 4(\pi N)^{-1} \\
& \times[f(s)+f(t)]v + (\pi N)^{-1}w + (4\pi^2N^2)^{-1}[F_N(s+t) + F_N(s-t) - 12N]v^2 \\
& + (4\pi^2N^2p^2)^{-1}[F_N(s+t) + F_N(s-t) - 2N(3 - p^2)]v^2 + (\pi Np^2)^{-1} \\
& \times \int_{-\pi}^{\pi} f(x)[f(x+s+t) + f(x+s-t)]\, dx \qquad (4.4) \\
& + (p\pi N)^{-1}\{[f(s)N^{-1}(F_N(s+t) + F_N(s-t)) - 2(3-p)(f(s) \\
& + f(t))]v - (2\pi N)^{-1}v^2[F_N(s+t) + F_N(s-t) - \underline{4N}(3-p)] - \frac{2-p}{2}
\end{aligned}$$

$$\times \int_{-\pi}^{\pi} f(x)[f(x+s+t) + f(x-s+t) + f(x+s-t) + f(x-s-t)]\, dx$$

$$+ 4\pi(2-p)f(s)f(t)\}$$

where for abbreviation we put

$$v = \int_{-\pi}^{\pi} f(x)\, dx, \qquad w = \int_{-\pi}^{\pi} f^2(x)\, dx$$

and $F_N(.)$ is the Fejer kernel. The underlined parts of (4.4) indicate the errors in the original formula of Scheinok.

By substituting (4.4) into (2.12), (2.17) and (2.23) and using delta function approximation to Fejer kernel, one can obtain the asymptotic error covariance of $\hat{\underset{\sim}{a}}'$ and $\hat{\sigma}^{2\,\prime}$. But the resulting formulas may be rather lengthy and complicated, so we only derive the expression for (2.12). After some simple calculations, we get

$$\begin{aligned} A_{k,i} &= N^{-1}v^2 a_k a_i(-9 + 15p^{-1} - 6p^{-2}) \\ &\quad + N^{-1}\sigma^2 r_{k-i} + 2N^{-1}\sigma^2\delta_{k,i}v(p^{-1} - 1) \\ &\quad + N^{-1}v^2 e_{k,i}(1 + p^{-2} - 2p^{-1}) \qquad (4.5) \\ &\quad + N^{-1}v^2 f_{k,i}(1 + p^{-2} - 2p^{-1}) + 2N^{-1}vg_{k,i} \\ &\quad \times(p^{-1} - 1) + (\pi N)^{-1}h_{k,i}(p^{-2} - p^{-1} + 0.5) \\ &\quad -(\pi N)^{-1}w_{k,i}(p^{-1} - 0.5) \end{aligned}$$

with

$$\begin{aligned} e_{k,i} &= (2\pi)^{-1} \int_{-\pi}^{\pi} B(s)B(-s)\, e^{j(k-i)s}\, ds \\ &= \sum_{n-n'=k-i} (-a_n)(-a_{n'}) \quad (n,n'=0,1,\ldots,m), \end{aligned} \qquad (4.6)$$

$$f_{k,i} = (2\pi)^{-1} \int_{-\pi}^{\pi} B^2(s)\, e^{j(k+i)s}\, ds$$
$$= \sum_{n+n'=k+i} (-a_n)(-a_{n'}), \tag{4.7}$$

$$g_{k,i} = (2\pi)^{-1} \int_{-\pi}^{\pi} B^2(s) f(s)\, e^{j(k+i)s}\, ds$$
$$= \sum_{n=0}^{m} \sum_{n'=0}^{m} (-a_n)(-a_{n'})\, r_{n+n'-k-i}, \tag{4.8}$$

$$h_{k,i} = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} B(s)B(t) \int_{-\pi}^{\pi} f(x)[f(x-s+t) + f(x-s-t)]\, dx$$
$$\times\, e^{j(ks+it)}\, ds\cdot dt, \tag{4.9}$$

$$w_{k,i} = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} B(s)B(t) \int_{-\pi}^{\pi} f(x)[f(x-s+t) + f(x-s-t)]\, dx$$
$$\times\, e^{j(ks+it)}\, ds\cdot dt. \tag{4.10}$$

In (4.9), by changing x to -x and using $f(x) = f(-x)$, we have $h_{k,i} = w_{k,i}$. Substitution of $f(x+s\pm t) = (2\pi)^{-1}\sum_{i=-\infty}^{\infty} r_i \exp[-ji(x+s\pm t)]$ into (4.9) yields

$$h_{k,i} = w_{k,i} = 2\pi \sum_{n=-\infty}^{\infty} r_n^2[(-a_{k-n})(-a_{i-n}) + (-a_{k-n})(-a_{i+n})] \tag{4.11}$$

where we define $-a_n = 0$ if $n > m$ or $n < 0$.

To check the validity of (4.5), we put p = 1, yielding the same result (2.27) in Chapter 2. Also to know the explicit value of (4.5), for example, let $\{x_t\}$ be a first-order AR process with $r_k - a^{|k|}$. In this case, it easily follows that $v = r_0 = 1$, $\sigma^2 = 1 - a^2$, $e_{1,1} = 1 + a^2$, $f_{1,1} = a^2$, $g_{1,1} = 0$ and $h_{1,1} = w_{1,1} = 2\pi\cdot 3a^2$. Therefore, we get

$$N\, E[(\Delta a)^2]_{miss} \approx p^{-2} + a^2(2p^{-2} - 3p^{-1}). \tag{4.12}$$

Since the missing rate is 1-p, the number of the net observations can be assumed to be Np with probability one as N tends to infinity. Thus it is reasonable to compare (4.12) with the error variance from the data of length Np without missing observations. The latter is

$$N\,E[(\Delta a)^2]_{cont} = p^{-1} - a^2p^{-1}. \qquad (4.13)$$

Hence, as long as $p < 1$ holds,

$$E[(\Delta a)^2]_{miss} > E[(\Delta a)^2]_{cont}. \qquad (4.14)$$

This inequality shows the serious effect of missing observations on estimation of parameters. This seriousness is increased as p tends to zero, since $E[(\Delta a)^2]_{miss}/E[(\Delta a)^2]_{cont} \approx (1+2a^2)p^{-1}/(1-a^2)$ for small p. Also, it is interesting to note that at $p = 2/3$, $N\,E[(\Delta a)^2] \approx p^{-2}$, independent of the system parameter a.

To conclude this section, the simulation result and the theoretical value (4.12) are compared. For $a = 0.5$ and $p = 0.5$, the former, calculated by averaging the squares of the estimation errors over 100 different data sets each of $N = 500$ length, is 4.478 while the latter is 4.5. The agreement is fairly good.

4.3 Regularly Missed Observations

Let $\{x_t\}$ be sampled in groups of α consecutive time instants separated β missed observations ($\alpha > \beta$). This situation may occur in the radar studies of the moon surface since during the reception of the radar echo, one must systematically cease the signal transmission so that there are time intervals without the reflections of the signals (Parzen 1963).

Let $\{d_t\}$ be defined similarly as in the previous section. Then, $\{d_t\}$ is a sequence of period $\alpha + \beta$. According to Jones (1962), we define the limit of the ratio of N, the total sample size, to the number of pairs available for estimating $r_k = E[x_t x_{t+k}]$ by

$$c_k = \lim_{N \to \infty} \frac{N}{\sum_{t=1}^{N-|k|} d_t \, d_{t+|k|}} . \tag{4.15}$$

Then, $c_k = c_{-k}$ and $\{c_k\}_{k=0}^{\infty}$ is also a sequence of period $\alpha+\beta$. The values of c_k during one period are as follows; $c_k = (\alpha + \beta)/(\alpha - k)$ for $0 \leqq k \leqq \beta$, $c_k = (\alpha + \beta)/(\alpha - \beta)$ for $\beta \leqq k \leqq \alpha$, and $c_k = (\alpha + \beta)/(k - \beta)$ for $\alpha \leqq k \leqq \alpha+\beta$. It is obvious that the consistent estimator for r_k is given by

$$\hat{r}''_k = \frac{1}{N} \sum_{t=1}^{N} c_k \, d_t \, d_{t+|k|} x_t \, x_{t+|k|} . \tag{4.16}$$

By defining the modified periodogram as

$$I''_N(s) = \frac{1}{2\pi N} \sum_{i=1}^{N} \sum_{k=1}^{N} d_i d_k c_{i-k} x_i x_k \, e^{-j(i-k)s} \tag{4.17}$$

, $\hat{r}''_k$ is expressed in terms of this modified periodogram as

$$\hat{r}''_k = \int_{-\pi}^{\pi} I''_N(s) \, e^{jks} \, ds. \tag{4.18}$$

From (4.17) it is easily shown that

$$E[\, I''_N(s) \,] \quad E[\, I_N(s) \,] + O(N^{-1}). \tag{4.18}$$

The term of order N^{-1} in (4.18) arises from the fact that $(c_k - N/\sum_{t=1}^{N-|k|} d_t d_{t+|k|})$ is of order N^{-1} but as is readily seen later, this term is of no importance for the ensuing analysis. Substitution of (4.16) into (2.5) and (2.14) gives the consistent estimators $\hat{\underset{\sim}{a}}''$, $\hat{\sigma}^{2}''$ for $\underset{\sim}{a}$, σ^2, respectively. Noting (4.18), the error covariance matrix of $\hat{\underset{\sim}{a}}''$ can be obtained by replacing $\mathrm{Cov}[I_N(s), I_N(t)]$ in (2.12) by $\mathrm{Cov}[I''_N(s), I''_N(t)]$. Hence, it is sufficient to know $\mathrm{Cov}[I''_N(s), I''_N(t)]$.

According to Jones (1962), $d_\nu d_\mu c_{\nu-\mu}$ in (4.17) is periodic so that it has the following two-dimensional representation;

$$d_\nu d_\mu c_{\nu-\mu} = \sum_{k,i} H_{k,i}\, e^{j(-\nu\lambda_k+\mu\lambda_i)} \tag{4.19}$$

where $\lambda_k \triangleq 2\pi k/(\alpha + \beta)$ and $k,i = -(\alpha + \beta - 1)/2,\ -(\alpha + \beta - 3)/2, \ldots, (\alpha + \beta - 1)/2$ if $\alpha + \beta$ is odd, or $-(\alpha + \beta - 2)/2,\ -(\alpha + \beta - 4)/2, \ldots, (\alpha + \beta)/2$ if $\alpha + \beta$ is even. In general, $H_{k,i}$ is very complicated as indicated in Jones(1962, page 458) but for $\beta = 1$ it reduces to $H_{k,k} = \delta_{k,0}$ and $H_{k,i} = (\alpha^{-1} - \delta_{k,0} - \delta_{i,0})/(\alpha - 1)$ $(k \neq i)$. Substituting (4.19) into (4.17) and introducing the discrete Fourier transform

$$J_N(s) \triangleq \sum_{t=1}^{N} x_t\, e^{-jts} \tag{4.20}$$

, $I''_N(s)$ is reexpressed as

$$I''_N(s) = \frac{1}{2\pi N} \sum_{k,i} H_{k,i}\, J_N(s+\lambda_k)\, J_N(-s-\lambda_i). \tag{4.21}$$

Since by the Gaussian assumption $J_N(s)$ is also Gaussian, it easily follows that

$$\mathrm{Cov}[I''_N(s), I''_N(t)] = \frac{1}{(2\pi N)^2} \sum_{k,i} \sum_{k',i'} H_{k,i}\, H_{k',i'} \{$$

$$E[J_N(s+\lambda_k) J_N(t+\lambda_{k'})] \cdot E[J_N(-s-\lambda_i) J_N(-t-\lambda_{i'})]$$

$$+ E[J_N(s+\lambda_k)J_N(-t-\lambda_{i'})]\cdot E[J_N(-s-\lambda_i)J_N(t+\lambda_{k'})]\} \quad . \qquad (4.22)$$

On the other hand, from Brillinger(1975, page 93) it follows that

$$E[\ J_N(s)J_N(t)\ ] = 2\pi\ f(s)\ D_N(s+t) + O(1) \qquad (4.23)$$

where $f(s)$ is the spectrum of $\{x_t\}$ and $D_N(.)$ is the Dirichlet kernel defined by

$$D_N(s) = \sum_{k=1}^{N} e^{-jks}. \qquad (4.24)$$

The second term of (4.23) is uniform in s, t so that its contributions to the integrals below can be neglected. From (4.22), (4.23) we have

$$\mathrm{Cov}[I''_N(s),I''_N(t)] = \frac{1}{N^2} \sum_{k,i,k',i'} H_{k,i}H_{k',i'}[f(s+\lambda_k)$$

$$\times f(t+\lambda_{i'})D_N(s+t+\lambda_k+\lambda_{k'})D_N(-s-t-\lambda_i-\lambda_{i'}) + f(s+\lambda_k)$$

$$\times f(t+\lambda_{k'})D_N(s-t+\lambda_k-\lambda_{i'})D_N(-s+t-\lambda_i+\lambda_{k'})]. \qquad (4.25)$$

Substituting this into (2.12) and using (4.24), the contribution of the first term in the square bracket of (4.25) to $A_{n,n'}$ in (2.12) is written as

$$\sum_{q=1}^{N} \sum_{q'=1}^{N} \int_{-\pi}^{\pi} B(s)f(s+\lambda_k)\ e^{-j(q-q'-n)s}ds \times \int_{-\pi}^{\pi} B(t)$$

$$\times f(t+\lambda_{i'})\ e^{-j(q-q'-n')t}\ dt\ e^{-j(\lambda_k+\lambda_{k'})q+j(\lambda_i+\lambda_{i'})q'} . \qquad (4.26)$$

From the definition of $B(s)$ and $f(s)$, we readily obtain

$$\int_{-\pi}^{\pi} B(s)f(s+\lambda_k)\ e^{-j(\nu-n)s}\ ds$$

$$= \sum_{i=0}^{m} (-a_i)r_{\nu-n+i}\ e^{j(\nu-n+i)\lambda_k} \triangleq \Theta_{\nu-n}(\lambda_k). \qquad (4.27)$$

Hence, (4.26) reduces to

$$\sum_{\nu=0}^{N-1} \Theta_{\nu-n}(\lambda_k)\Theta_{\nu-n'}(\lambda_{i'})\, e^{-j(\lambda_k+\lambda_{k'})\nu} \sum_{q=1}^{N-\nu} e^{-j(\lambda_k+\lambda_{k'}-\lambda_i-\lambda_{i'})q}$$

$$+ \sum_{\nu=-N+1}^{-1} \Theta_{\nu-n}(\lambda_k)\Theta_{\nu-n'}(\lambda_{i'})\, e^{-j(\lambda_k+\lambda_{k'})\nu} \sum_{q=1-\nu}^{N} e^{-j(\lambda_k+\lambda_{k'}-\lambda_i-\lambda_{i'})q}$$

$$= \sum_{\nu=-\infty}^{\infty} \Theta_{\nu-n}(\lambda_k)\Theta_{\nu-n'}(\lambda_{i'})\, e^{-j(\lambda_k+\lambda_{k'})\nu} D_N(\lambda_k+\lambda_{k'}-\lambda_i-\lambda_{i'}) + O(1) \tag{4.28}$$

where we use the fact that $|\Theta(\lambda_k)|$ is exponentially decreasing.

It is well known that

$$D_N(s) = \begin{cases} N & \text{for } s \equiv 0 \ (\bmod 2\pi) \\ O(1) & \text{otherwise.} \end{cases} \tag{4.29}$$

Thus, the value of (4.28) is of order N if and only if $\lambda_k+\lambda_{k'}-\lambda_i -\lambda_{i'} \equiv 0 \ (\bmod 2\pi)$. Since $-\pi < \lambda_k < \pi$ if $\alpha+\beta$ is odd and $-\pi < \lambda_k \leqq \pi$ if $\alpha+\beta$ is even, it follows that $|\lambda_k+\lambda_{k'}-\lambda_i-\lambda_{i'}| < 4\pi$ regardless of the parity of $\alpha+\beta$. Hence, from the definition of λ_k, the above possibilities are $k+k'-i-i' = 0, \pm(\alpha+\beta)$.

In a similar way, the integral due to the second term in the square bracket in (4.25) is calculated as

$$\sum_{\nu=-\infty}^{\infty} \Theta_{\nu-n}(\lambda_k)\Theta_{-\nu-n'}(\lambda_{k'})\, e^{-j(\lambda_k-\lambda_{i'})\nu} D_N(-\lambda_i+\lambda_{k'}+\lambda_k-\lambda_{i'}) + O(1). \tag{4.30}$$

This is also of order N if and only if $k+k'-i-i' = 0, \pm(\alpha+\beta)$.

Thus, the asymptotic value of $A_{n,n'}$ is given by

$$N^{-1} \sum_{\substack{k,i,k',i' \\ k+k'-i-i'=0,\pm(\alpha+\beta)}} H_{k,i}H_{k',i'} \{ \sum_{\nu=-\infty}^{\infty} \Theta_{\nu-n}(\lambda_k)\Theta_{\nu-n'}(\lambda_{i'})$$

$$\times\, e^{-j(\lambda_k+\lambda_{k'})} \quad + \theta_{\nu-n}(\lambda_k)\theta_{-\nu-n'}(\lambda_{k'})\, e^{-j(\lambda_k-\lambda_{i'})}\}. \quad (4.31)$$

However, (4.31) is too complicated to understand the property of the estimator, so we consider a simple example. Let $\{x_t\}$ be a first order AR process with $r_k = a^{|k|}$ ($|a| < 1$). For $n = n' = 1$, (4.27) is

$$\theta_{\nu-1}(\lambda) \quad = \quad (r_{\nu-1}\, e^{-j\lambda} - a\, r_\nu)\, e^{j\nu\lambda}.$$

From this, the infinite summation in (4.31) is given by

$$(z_k - a^2)(z_{i'} - a^2)\frac{z_{k'-i'}}{1 - a^2 z_{k'-i'}} + (z_k - 1)(z_{i'} - 1)\times$$

$$\frac{a^2}{1 - a^2 z_{i'-k'}} + (z_k - a^2)(z_{k'} - 1)\frac{a^2 z_{k'-i'}}{1 - a^2 z_{k'-i'}}$$

$$+ a^2(z_k - 1)(z_{k'} - 1) + (z_k - 1)(z_{k'} - a^2)\frac{a^2 z_{i'-k'}}{1 - a^2 z_{i'-k'}}$$

with $z_k \triangleq \exp(-j\lambda_k)$. Numerical calculations were performed for various values of a, α with $\beta = 1$. To see the effect of missing observations, we compare $N\, E[(\Delta a)^2]_{miss}$ with the error variance from the data of length 2N/3 without missing observations, since for $\alpha = 2$, $\beta = 1$, the number of the net observations is 2N/3. The latter is

$$\frac{2}{3} N\, E[(\Delta a)^2]_{cont} \quad \simeq \quad 1 - a^2 \qquad (4.32)$$

Table 4.1 shows these values for $|a| = 0.1, 0.2, \ldots, 0.9$. It is interesting to note that the correlation of the data becomes strong , that is, for $|a| > 0.8$, the degrading effect of regularly missed observations disappears. This is quite different from the case of randomly missed observations, since from (4.12) and (4.13) we have $E[(\Delta a)^2]_{miss}/E[(\Delta a)^2]_{cont} = (p^{-1}+a^2(2p^{-1}-3))/(1-a^2) = 1.5/(1-a^2)$ with missing rate $1-p = 1/3$, so the relative degrading effect becomes more and more serious as $|a| \to 1$. Table 4.2 shows the behavior of $N\,E[(\Delta a)^2]_{miss}$ for increasing values of α with $\beta = 1$ fixed. The convergence to $1-a^2$ is apparent but the converging rates are fairly different. That is, for small $|a|$, the rate is high whereas for larger $|a|$ near 1, the convergence is considerably slow. This phenomenon also occurs in the case of randomly missed observations and can be explained as follows.

$\|a\|$	$NE[(\Delta a)^2]_{miss.}$	$NE[(\Delta a)^2]_{cont.}$
0.1	2.940	1.485
0.2	2.765	1.440
0.3	2.486	1.365
0.4	2.126	1.250
0.5	1.714	1.125
0.6	1.289	0.960
0.7	0.884	0.765
0.8	0.527	0.540
0.9	0.231	0.285

Table 4.1 Comparison of the variances with and without missing observations.

$\|a\|$	$\alpha=10$	$\alpha=20$	$\alpha=30$	$\alpha=\infty$
0.1	1.222	1.093	1.053	0.990
0.2	1.181	1.063	1.022	0.960
0.3	1.129	0.014	0.978	0.910
0.4	1.055	0.944	0.908	0.840
0.5	0.959	0.853	0.818	0.750
0.6	0.839	0.740	0.707	0.640
0.7	0.689	0.607	0.575	0.510
0.8	0.501	0.447	0.421	0.360
0.9	0.266	0.252	0.239	0.190

Table 4.2 Convergent behavior of $N \cdot E[(\Delta a)^2]_{miss}$ for increasing α.

Since the covariance estimator (4.16) is based on filling the missed observations with zeros, *a priori* mean, this estimate does not make any use of the information about the data correlation. Thus, the degrading effect vanishes promptly as $\alpha \to \infty$ in the white noise case whereas it is still non-neglegible for the data with strong correlation. Also we can note a quite curious phenomenon in Table 4.2. That is, at $a = 0.9$, the variance for $\alpha = 2$ is smaller than those for $\alpha = 10, 20, 30$! At present there is no explanation for this counter-intuitive result. Perhaps, this is due to the suboptimality of the present estimation procedure and the maximum likelihood estimate may not possess such a property.

To see the validity of the theoretical results, in Table 4.3, we present simulation results where empirical variances were obtained by averaging squares of estimation errors over M sets of data each of length N. We can see a fairly good agreement between the theoretical and experimental results.

a	Number of data N	Number of data sets M	$N \cdot E[(\Delta a)^2]_{miss}$ by Theory	By Simulations
0.8	1000	500	$\alpha = 2$ 0.527 $\alpha = 10$ 0.501	$\alpha = 2$ 0.576 $\alpha - 10$ 0.493
0.9	1000	500	$\alpha = 2$ 0.231 $\alpha = 10$ 0.266 $\alpha = \infty$ 0.190	$\alpha = 2$ 0.262 $\alpha = 10$ 0.277 $\alpha = \infty$ 0.202
0.9	1000	900	$\alpha = 2$ 0.231 $\alpha = 10$ 0.266	$\alpha = 2$ 0.262 $\alpha = 10$ 0.282
0.95	1000	500	$\alpha \quad 2$ 0.108 $\alpha = 10$ 0.134	$\alpha = 2$ 0.134 $\alpha = 10$ 0.147

Table 4.3 The simulation results to show the validity of the theoretical analysis.

4.4 Conclusion

We have derived the error covariance matrices of the estimates of the AR parameters based on randomly and regularly missed observations by using the periodogram technique. In these cases, this method is particularly powerful since most of other conventional techniques, such as the calculation of Fisher information matrices , break down. By this method, we can use much of the classical works about the variously modified periodograms to modern parametric problems.

At first sight we are apt to think negative effects of missing observations. But from the results of Section 4.3, in some cases, we can positively utilize the concept of missing observations to improve the performance of the estimate if the number of the observations is fixed but time instances at which the observations are made can be changed. For example, for a first order AR process with $a = 0.9$, about 20 % reduction of the variance is gained if we allocate the total observations of length, say, $N = 500$ over 750 instances to form regularly missed observations with $\alpha = 2$, $\beta = 1$. This is the first example after Neave (1970) which positively utilizes missing observations.

Chapter 5

Estimation of Frequencies of Sinusoids in Noise

5.1 Introduction

In recent years, the autoregressive (AR) spectral analysis method proposed by Burg (1967) and Pazen (1970) has received much attention and has been frequently used in many fields of science. This is because the AR method has been recognized to have much greater resolution than the traditional methods, such as the lag window technique. As for the theoretical aspects of this method, Lacoss (1971) studied the case where time series are made up of several sinusoids and white noise. But the statistical properties of this method have not been well studied. Recently, Baggeroer (1976) introduced a particular assumption about the data structure which enables us to apply the theory of multivariate analysis to this AR (MEM) method and derived many interesting statistical properties. In a sense, his result can be considered as a generalization of Akaike's earlier work (Akaike 1969 b) where the data are assumed to be generated by a pure finite-order AR process.

In the next section of this chapter, by using the periodogram technique, the statistical analysis of the AR method applied to the data consisting of several sinusoids plus stationary noise is performed with emphasis on its asymptotics in a different way from Baggeroer (1976). At first, we note that the AR spectral estimator

(ARSPE) is asymptotically equivalent to a smoothed periodogram with a data-dependent "spectral window". From this expression, the variances of the estimator can be determined and their behavior turns out to be similar to Kromer and Berk's earlier results for stationary processes (Kromer 1969, Berk 1974). Next, we consider the frequency measurement aspect of the AR method. Since it is reasonable to expect that the AR spectrum has several sharp peaks at the corresponding sinusoidal frequencies with sufficiently high order, the frequency at which the sample AR spectrum gives a sharp peak can be used as a frequency estimate. Under the assumption that the estimation error is small, the asymptotic variance of that estimator is derived and compared with that of the conventional Fourier method and recent results of Lang (1979).

We note the dependence of the variance on the data length and the signal-to-noise ratio (SNR). They play a kind of dual role in the expression of the variance at different SNR regions.

Next, we investigate statistical properties of Pisarenko frequency estimator (Pisarenko 1973) which is essentially regarded as an eigenvalue and eigenvector analysis of the covariance matrix. The functional dependence of the variance of the Pisarenko estimator on the data length and SNR is same with the AR method. We discuss merits and demerits of these modern methods over the conventional Fourier method.

5.2 Performance Analysis for the AR Method

5.2.1 General Analysis

We assume that the time series under consideration consist of p sinusoids and a zero-mean stationary Gaussian noise n_t, that is, $\{x_t\}$ is expressed as

$$x_t = \sum_{i=1}^{p} [A_i e^{j\omega_i t} + A_i^* e^{-j\omega_i t}] + n_t \qquad (5.1)$$

where " $*$ " denotes the complex conjugate. Hence, the autocovariance function r_k is

$$r_k = \sum_{i=1}^{p} 2|A_i|^2 \cos(\omega_i k) + q_k \qquad (5.2)$$

where $q_k = E[n_t n_{t+k}]$. In (5.2) r_k is defined by the limit of $\hat{r}_k$ in (2.4) as $N \to \infty$, and q_k is written as above because of the ergodicity of $\{n_t\}$. As an alternative definition, we may assume that A_i is expressed as $B_i \exp(j\phi_i)$ where ϕ_i is a random phase angle distributed uniformly over $[0,2\pi]$. Then r_k can be defined as the ensemble average $E[x_t x_{t+k}]$. However, these two definitions lead the same autocovariance function and, as is easily seen later, the same statistical properties of the AR method, so that we assume the first definition to be the following.

Given a set of data $\{x_1, x_2, \ldots, x_N\}$, it is well known that that the AR estimate is calculated in the same way as in Sections 2.2 and 2.3 *despite* $\{x_t\}$ under consideration is *not* a pure AR process. We call the resulting estimated spectrum $\hat{f}(s)$ as the estimate of the AR spectrum $f(s)$, or the AR spectral estimator (ARSPE).

Of course, $f(s)$ is different from the true spectrum $W(s)$ in (A-5.9) of $\{x_t\}$, but with sufficient high order m, the shape of $f(s)$ closely resembles that of $W(s)$. Lacoss (1971) studied this problem in detail. From (2.31), (2.33), (2.10), and (A-2.1), the estimation error $\Delta f(s) = \hat{f}(s) - f(s)$ is asymptotically expressed as

$$\Delta f(s) \quad \int_{-\pi}^{\pi} f(s)B(t)[K(s,t) + \sigma^{-2}B(-t)]\ I_N(t)\ dt - f(s) \tag{5.3}$$

where $K(s,t)$ is defined by (2.35). This shows that

$$\hat{f}(s) \approx \int_{-\pi}^{\pi} G(s,t)I_N(t)\ dt \tag{5.4}$$

with

$$\begin{aligned} G(s,t) &= f(s)B(t)[K(s,t) + \sigma^{-2}B(-t)] \\ &\triangleq f(s)g(s,t). \end{aligned} \tag{5.5}$$

Thus, the ARSPE can be viewed as a smoothed periodogram with a data-dependent "spectral window" $G(s,t)$. Obviously, this indicates the data-adaptivity of the AR method which was mentioned but not explicitly shown by Lacoss (1971). Needless to say, in the classical Blackman-Tukey method, $G(s,t)$ is invariant, ie, $G(s,t) = w(s-t)$ and moreover, $w(.)$ does not contain any parameters of $\{x_t\}$. From (5.3), it readily follows that

$$\begin{aligned} E\left[\frac{\Delta f(\mu)}{f(\mu)} \cdot \frac{\Delta f(\nu)}{f(\nu)}\right] &\approx \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} g(\mu,s)g(\nu,t)\mathrm{Cov}[I_N(s),I_N(t)]\ ds\cdot dt \\ &+ \left\{\int_{-\pi}^{\pi} g(\mu,s)E[I_N(s)]ds - 1\right\}\left\{\int_{-\pi}^{\pi} g(\nu,t)E[I_N(t)] - 1\right\} \end{aligned} \tag{5.6}$$

From Appendix 5.1, the second term of (5.6) is of order N^{-2} and can be neglected. Applying the result of Appendix 5.2, the remaining term becomes

$$\frac{2\pi}{N}\{\sum_{k=1}^{p} 2|A_k|^2[g(\mu,\omega_k)g(\nu,-\omega_k) + g(\mu,-\omega_k)g(\nu,\omega_k)$$

$$+ g(\mu,\omega_k)g(\nu,\omega_k) + g(\mu,-\omega_k)g(\nu,-\omega_k)]Q(\omega_k)$$

$$+ \int_{-\pi}^{\pi}[g(\mu,s)g(\nu,s) + g(\mu,s)g(\nu,-s)]Q^2(s)\,ds\}. \quad (5.7)$$

At present, the expression (5.7) does not permit clear-cut analytic interpretation, so that we must resort to numerical calculations to examine its behavior. This will be done in the next subsection.

Next, we consider the statistical fluctuation of the estimator $\hat{\omega}$ for the frequency ω at which $f(s)$ has a peak value. Since $f(s)$ and $\hat{f}(s)$ take extreme values at ω and $\hat{\omega}$, respectively, it follows that

$$\frac{d}{ds}\, B(s)B(-s)|_{s=\omega} = \frac{d}{ds}\, \hat{B}(s)\hat{B}(-s)|_{s=\hat{\omega}} = 0$$

or

$$B'(\omega)B(-\omega) + B(\omega)B'(-\omega) = 0$$

$$\hat{B}'(\hat{\omega})\hat{B}(-\hat{\omega}) + \hat{B}(\hat{\omega})\hat{B}'(-\hat{\omega}) = 0 \quad (5.8)$$

where " ' " denotes the differential operation with respect to s and

$$\hat{B}(s) = \sum_{i=0}^{m} (-\hat{a}_i)\, e^{-jis}. \quad (5.9)$$

As above, the error $\Delta\omega = \hat{\omega} - \omega$ can be assumed to small for sufficiently large N, so that the following approximations are valid:

$$\hat{B}'(\hat{\omega}) \approx \hat{B}'(\omega) + \hat{B}''(\omega)\Delta\omega ,$$

$$\hat{B}(-\hat{\omega}) \simeq \hat{B}(-\omega) + \hat{B}'(-\omega)\Delta\omega.$$

Also, from (5.9)

$$\hat{B}(s) = B(s) - \Delta\underset{\sim}{a}^T \underset{\sim}{E}(s)$$

with $\underset{\sim}{E}^T(s) = [e^{-js}, \ldots, e^{-jms}]$. Substituting these into (5.8) and neglecting higher order terms like $\Delta a^T \Delta\omega$, we have

$$\Delta\omega \simeq \Delta\underset{\sim}{a}^T \frac{\underset{\sim}{E}(-\omega)B'(\omega) + \underset{\sim}{E}(\omega)B'(-\omega) + \underset{\sim}{E}'(\omega)B(-\omega) + \underset{\sim}{E}'(-\omega)B(\omega)}{2B'(-\omega)B'(\omega) + B''(\omega)B(-\omega) + B''(-\omega)B(\omega)}$$

$$= \Delta\underset{\sim}{a}^T \underset{\sim}{C}(\omega). \qquad (5.10)$$

Defining $\underset{\sim}{R}^{-1}\underset{\sim}{C}(\omega) = [z_1(\omega), \ldots, z_m(\omega)]^T$ and using (2.10) give

$$\Delta\omega \simeq \int_{-\pi}^{\pi} B(s)Z(\omega,s)I_N(s)\, ds$$

with

$$Z(\omega,s) = \sum_{k=1}^{m} z_k(\omega)\, e^{jks}.$$

Hence, the variance $E[(\Delta\omega)^2]$ is calculated by applying the formula (A-5.27) in Appendix 5.2 with

$$F(s,t) = B(s)Z(\omega,s)B(t)Z(\omega,t). \qquad (5.11)$$

It should be noted that the variance is of order N^{-1}.

5.2.2 Numerical Examples

Although we have derived the fundamental formulas for evaluating the statistical properties, they do not allow simple analytical interpretations. Thus, numerical studies were carried out for the case where the data are made up of a single sinusoid and white noise, that is,

$$x_t = A e^{j\ t} + A^* e^{-j\ t} + n_t.$$

It is obvious from the above development that (5.7) and $N \cdot E[(\Delta\omega)^2]$ depend explicitly on the SNR $P = 2|A|^2/q_0$. Thus, without loss of generality, we assume that $2|A|^2 = P$ and $q_k = \delta_{k,0}$ $(Q(s) = 1/2\pi)$.

$S = \pi k/20$	Asymptotic variances		Empirical variances
	Total	Four terms	
k = 0	62.452	0.043	67.875
1	23.932	0.006	24.191
2	29.643	0.034	35.241
3	33.021	0.302	40.570
4	18.616	0.048	17.825
5	41.006	29.324	45.406
6	21.211	0.003	23.183
7	30.238	0.547	30.770
8	32.042	0.177	32.026
9	30.983	0.200	31.729
10	30.151	0.043	31.948
11	30.397	0.089	37.778
⋮	⋮	⋮	⋮
17	33.267	0.017	30.031
18	29.407	0.001	31.636
19	25.062	0.023	23.515
20	61.885	0.043	58.476

Table 5.1 Asymptotic and empirical variances of the ARSPE for a sinusoid plus white noise.

The first column of Table 5.1 shows the numerical values of $N \cdot E[(\Delta f(\nu)/f(\nu))^2]$ for $\omega = \pi/4$, $m = 15$, and $P = -10$ dB. The second column shows the contributions due to the first four terms of (5.7), i.e., the summation part. It can be seen that except at $\nu = \omega$, these are negligibly small. To see the validity of the asymptotic expression (5.7), in the third column, we also present the simulation results which were obtained by averaging the squares of the estimation error $\Delta f(s)$ over 100 different data sets each of $N = 1000$ length. The agreements are fairly good. From this table it can be inferred that, except at frequencies near ω, the asymptotic variances are roughly approximated as

$$N \cdot E[(\frac{\Delta f(\nu)}{f(\nu)})^2] \doteq \begin{cases} 2m & \text{for } \nu \neq 0,\pi \\ 4m & \text{for } \nu = 0,\pi. \end{cases} \qquad (5.12)$$

Although we do not present detailed numerical tables here, the relation (5.12) holds for other cases such as multiple sinusoids plus nonwhite noise. However, the analytic proof of (5.12) is not known at present. It should be pointed out that Kromer (1969) and Berk (1974) have obtained the same result with (5.12) for the AR spectral estimate applied to a fairly wide class of stationary processes. In this respect, see also Section 2.3.

Now let us examine the variance of the main peak frequency fluctuation . Here $\Delta\omega$ is interpreted as the difference between the main peak frequencies of $f(s)$ and $\hat{f}(s)$. To see the explicit dependence of $N \cdot E[(\Delta\omega)^2]$ on P, we briefly analyze the case where $P << 1$. From (5.2) and (2.3), $\underset{\sim}{a}$ is of order P. Thus, $B(s) = 1 - O(P)$, $B'(s)$ $O(P)$, and $B''(s) = O(P)$ where $O(P)$ means that this is of order P.

But from (5.10), $\underset{\sim}{C}(\omega)$ becomes $O(P^{-1})$, and from (5.11), $F(s,t) = O(P^{-2})$. Consequently, we obtain

$$E[(\Delta\omega)^2] \doteq \alpha \frac{P^{-2}}{N} \quad \text{for } P \ll 1 \qquad (5.13)$$

where α is a constant depending on ω and m. Since it is unknown whether (5.13) holds for intermidiate and large P, the numerical calculations were performed. Fig. 5.1 shows the values of $N \cdot E[(\Delta\omega)^2]$ with $\omega = \pi/4$ and m = 5, 10, 15 versus P.

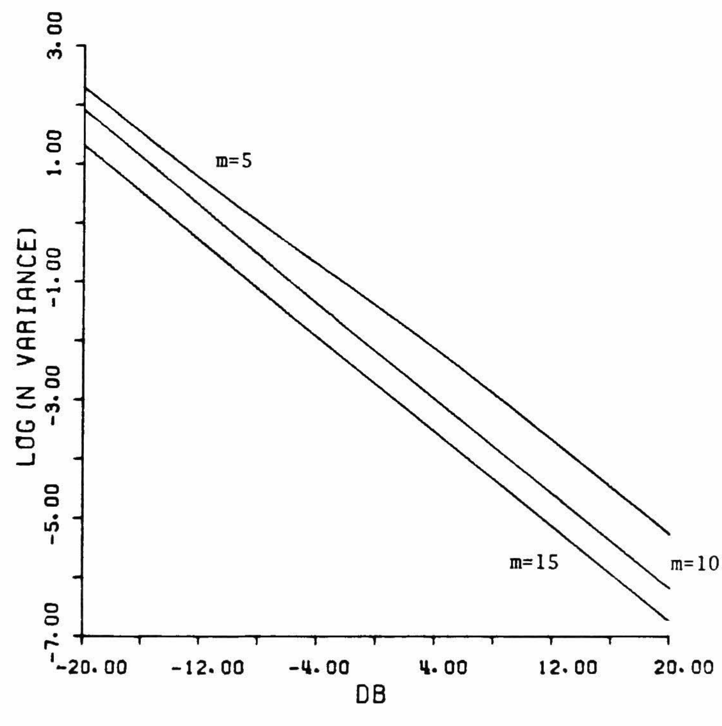

Fig. 5.1 The asymptotic variances of the AR frequency estimator with various autoregression orders.

The vertical axis is logarithmically scaled. Since the graphs for $m = 10, 15$ are almost straight lines and their inclinations are almost -0.2, it can be deduced that (5.13) approximately holds at all the range of P. The next thing that can be seen from Fig. 5.1 is the dependency on the length of autoregression, i.e., m. The gains obtained by increasing m by 5 are about 7 to 8 dB. At first sight, this observation seems to be against our intuition since, in many statistical problems, increasing the number of parameters causes larger statistical variations in final estimators. To explain the above fact, the shapes of $f(s)$ for $m = 5, 15$ are plotted in Figs.5.2 and 5.3 where solid lines indicate $f(s)$ and square marks show the sample spectra $\hat{f}(s)$.

It is intuitively clear that the more the shape is sharp, or in other words, the " bandwidth " of $f(s)$ is narrow, the more the estimation of ω is easy. In this case, the ultimate performance is more affected by the narrower bandwidth than the larger statistical variation. If the condition $Pm \gg 1$ is met, as shown in Lacoss (1971), $f(s)$ has a main peak at $s = \omega$. Hence, the above-mentioned $\hat{\omega}$ becomes an unbiased estimator for ω. From Figs.5.2 and 5.3, this unbiasedness holds for $P = -10$ dB, $m = 15$, but not for $P = -10$ dB , $m = 5$.

Lastly, we briefly discuss the implication of the assumption that $\Delta\omega$ is small. From Figs. 5.2 and 5.3, we see that for $N = 1000$, the sample spectra are close to the true spectra $f(s)$, so the assumption is valid. However, when N is reduced to 200, the shapes of $\hat{f}(s)$ are quite different from $f(s)$, and the spurious peaks have emerged. (See Figs. 5.4 and 5.5).

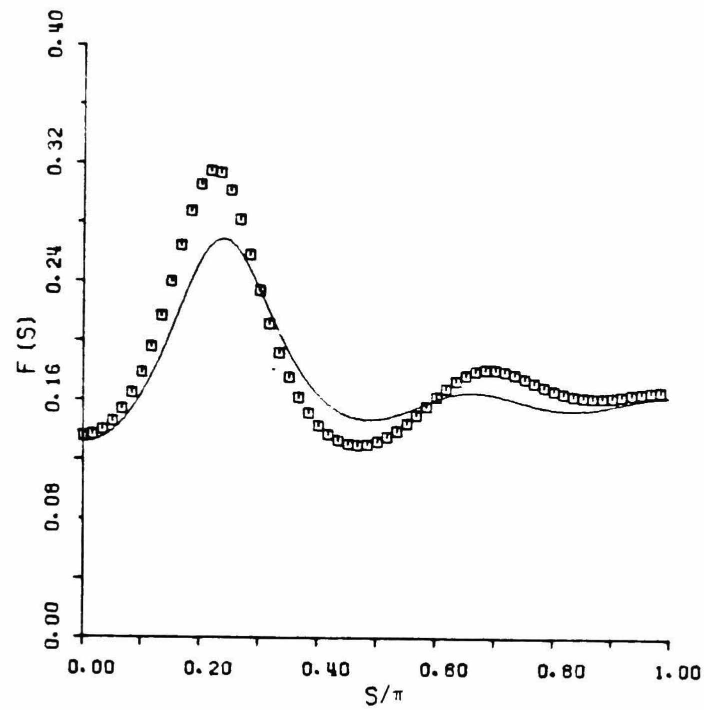

Fig 5.2 True and sample AR spectra for m = 5, P = -10 dB, and N 1000.

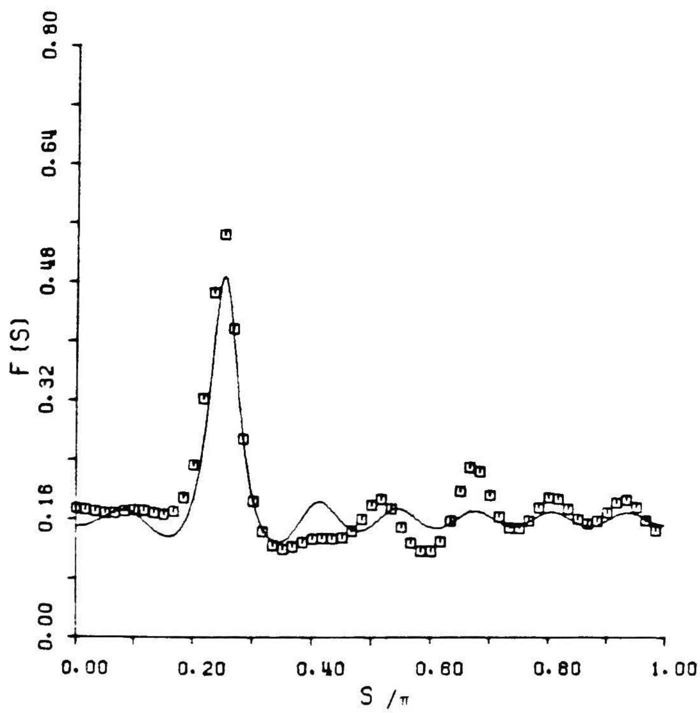

Fig. 5.3 True and sample AR spectra for m = 15, P = -10 dB, and N = 1000.

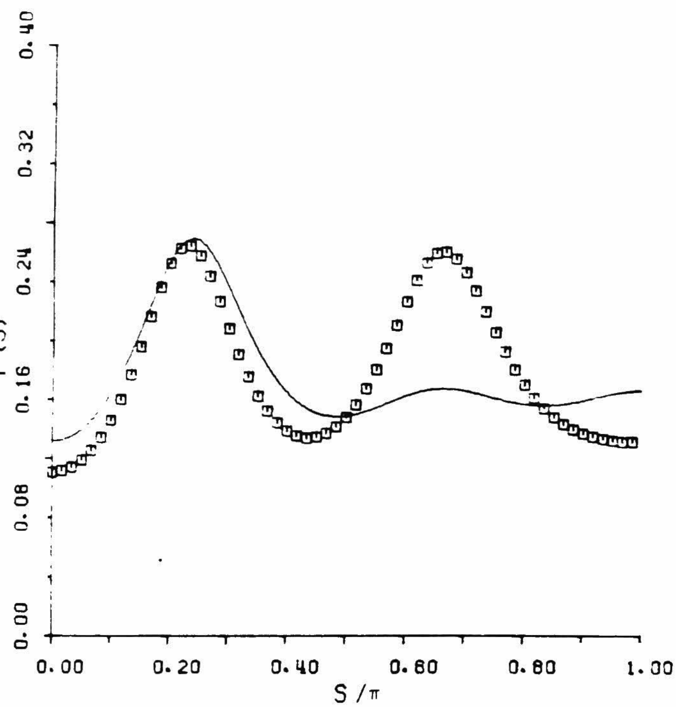

Fig. 5.4 An ill-behaved sample AR spectrum for m =-5, P=-10 dB, N= 200.

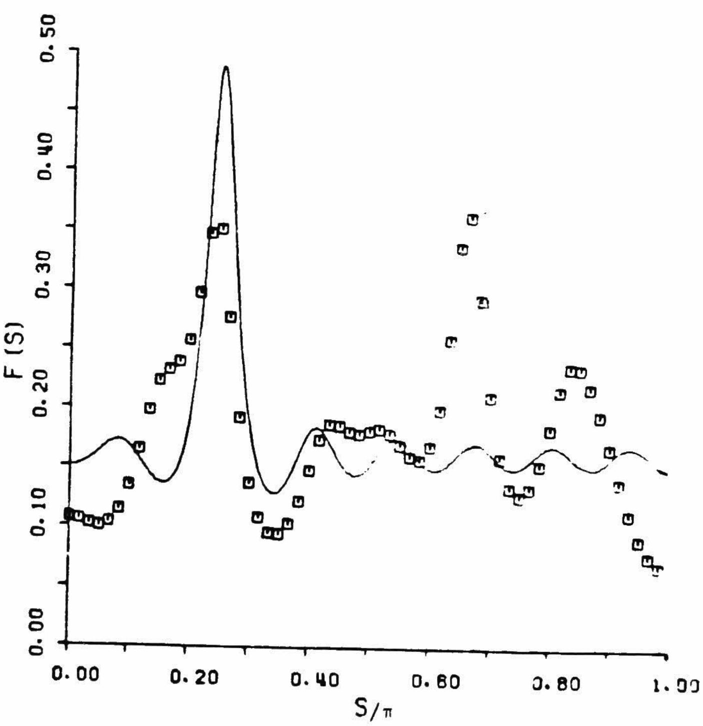

Fig. 5.5 An ill-behaved sample AR spectrum for m=15, P=-10 dB, N= 200.

Hence, taking these spurious peak frequencies as $\hat{\omega}$ causes large estimation error $\Delta\omega$, thereby breaking down our "local" analysis. This pheneomenon can be considered as a kind of "threshold effect" inherent in any nonlinear processing of data. Thus, (5.13) is valid only for $N \gtrsim N_{thr}(P)$ where $N_{thr}(P)$ is a function of P and specifies the threshold. However, its determination seems to be quite difficult by our asymptotic analysis.

5.2.3 Discussion

As another freqency estimation method, Fourier analysis has been widely used. The procedure is simply finding the frequency at which the periodogram takes an extreme value. If the quadrature component of (5.1) is available, according to Rife & Boorstyn (1974) , this method coincides with the maximum likelihood (ML) method.

It was shown that if the SNR is sufficiently high, or above the threshold, the Cramér-Rao bound (eq.(17) in Rife & Boorstyn 1974) is almost met by the ML frequency estimator; that is, the variance is inversely proportional to the SNR P and the cube of the data length N. But this implies that (5.13) is no longer true for P >> 1, since in this region, there is a possibility that (5.13) becomes smaller than the corresponding Cramér-Rao bound, an obvious contradiction. So we *must* remark that plots of the simulated variances of the AR frequency estimator will *rise up* from the lines in Fig. 5.1 at certain values of P. To obtain the theoretical curves for high SNR region, an entirely different analysis technique is required. Actually, Lang (1979) has just discussed this problem. He showed that using covariance and modified covariance methods for estimating the AR parameters, the variances at high SNR region are given by

$$E[(\Delta\omega)^2] = \begin{cases} \dfrac{8}{3m(N-m)^2P} & \text{for covariance method} \\ \dfrac{2}{m(N-m)^2P} & \text{for modified covariance method.} \end{cases} \tag{5.14}$$

Thus taking $m_{opt} = N/3$ makes (5.14) minimum. The corresponding Cramér-Rao bound is $12/N^3P$. The order-optimized covariance and modified covariance methods produce the performances 3/2 and 9/8 greater than the Cramér-Rao bound, respectively. This shows nearly optimum nature of the above two methods. But it is well known that the fitting order m must be taken as smaller order than $\sqrt{N}$ to make the estimates consistent. The above m_{opt} obviously violates this this condition. Thus, Lang's result only applies for the case where the number of data N is small while the SNR P is quite high as he assumed. In this case, the consistency about N is tactly replaced by the consistency *in P.*

However, in practice, such a case seldom occurs. If N is sufficiently large and P is moderately large, the above two methods coincide with the AR method with *finite* m. Then, from (5.14), the variance of the AR method is proportional to $N^{-2}P^{-1}$ since m << N. Thus we see a kind of dual relation in the expression of $E[(\Delta\omega)^2]$ about N and P.

5.3 Performance Analysis for the Pisarenko Method

5.3.1 Introduction

Pisarenko (1973) proposed a new technique for estimation of frequencies of sinusoids in white noise by examining the minimum eigenvalue and the corresponding eigenvector of the covariance matrix of observations. Frost (1977) illustrated the effectiveness of this method with several numerical examples, but as Baggeroer (1978) pointed out, statistical properties of this method have not been fully understood. We devote this section to investigate some statistical aspects of the method using the same technique as above. We also give a computationally simple version of the Pisarenko method. This supplies an estimation-theoretic interpretation for the results of Cantoni & Butler (1976).

5.3.2 The Pisarenko Method and Its Modification

We assume that the time series under consideration are given by (5.1) with Gaussian white noise sequence $\{n_t\}$, that is,

$$s_t = \sum_{i=1}^{p} [A_i e^{j\omega_i t} + A_i^* e^{-j\omega_i t}] \tag{5.15}$$

$$x_t = s_t + n_t$$

with $E[n_t n_{t'}] = \sigma^2 \delta_{t,t'}$. It is obvious that the signal $\{s_t\}$ satisfies the following 2p-th order linear difference equation

$$s_t = \sum_{k=1}^{2p} a_k s_{t-k} \tag{5.16}$$

where the equation

$$1 - a_1 z^{-1} - \dots - a_{2p} z^{-2p} = 0 \tag{5.17}$$

has the roots $z = \exp(\pm j\omega_i)$ ($i=1,\ldots, p$), or $B(\pm\omega_i) = 0$ where $B(s)$ is defined by (2.11) with $m = 2p$. From this, it immediately follows the symmetric property;

$$\begin{aligned} a_{2p} &= -1 \\ a_i &= a_{2p-i} \ (i=1,\ldots,p-1). \end{aligned} \tag{5.18}$$

Substituting $s_t = x_t - n_t$ into (5.16) shows that $\{x_t\}$ is expressed as a special ARMA(2p,2p) process

$$x_t - \sum_{i=1}^{2p} a_i x_{t-i} = n_t - \sum_{i=1}^{2p} a_i n_{t-i}. \tag{5.19}$$

Defining

$$\begin{aligned} \underset{\sim}{\alpha}^T &= [1 \ -a_1 \ \ldots \ -a_{2p}] \\ \underset{\sim}{x}^T &= [x_t \ x_{t-1} \ldots \ x_{t-2p}] \\ \underset{\sim}{n}^T &= [n_t \ n_{t-1} \ldots \ n_{t-2p}] \\ \underset{\sim}{s}^T &= [s_t \ s_{t-1} \ldots \ s_{t-2p}] \end{aligned} \tag{5.20}$$

, then (5.19) is $\underset{\sim}{x}_t^T \underset{\sim}{\alpha} = \underset{\sim}{n}_t^T$. Thus, $E[\underset{\sim}{x}_t \underset{\sim}{x}_t^T] \underset{\sim}{\alpha} = E[\underset{\sim}{x}_t \underset{\sim}{n}_t^T]\underset{\sim}{\alpha} = E[(\underset{\sim}{s}_t + \underset{\sim}{n}_t) \underset{\sim}{n}_t^T]\underset{\sim}{\alpha} = E[\underset{\sim}{n}_t \underset{\sim}{n}_t^T]\underset{\sim}{\alpha} = \sigma^2\underset{\sim}{\alpha}$ since $E[s_t n_{t'}] = 0$ for all t, t'. That is,

$$\underset{\sim}{C} \underset{\sim}{\alpha} = \sigma^2 \underset{\sim}{\alpha} \tag{5.21}$$

where $\underset{\sim}{C}$ is the $(2p+1)\times(2p+1)$ covariance matrix such that $(\underset{\sim}{C})_{ik} = r_{i-k}$ with $r_k = E[x_t x_{t+k}]$. Since $\underset{\sim}{C} = E[\underset{\sim}{s}_t \underset{\sim}{s}_t^T] + \underset{\sim}{I}_{2p}$ ($\underset{\sim}{I}_n$ denotes the n×n unit matrix.), (5.21) implies that σ^2 is the minimum eigenvalue of $\underset{\sim}{C}$ and $\underset{\sim}{\alpha}$ is the corresponding eigenvector. The Pisarenko estimate is obtained by replacing r_k in (5.21) by its estimate $\hat{r}_k$ in (2.4). The frequencies ω_i's are estimated by the same method as in Section 5.2. Now we state a modified estimate by noting the relation (5.18). Define

$$\tilde{\underset{\sim}{x}}_t^T = [y_t+y_{t-2p},\ y_{t-1}+y_{t-2p+1},\ \ldots,\ y_{t-p+1}+y_{t-p-1},\ by_{t-p}]$$

$$\tilde{\underset{\sim}{n}}_t^T = [n_t+n_{t-2p},\ n_{t-1}+n_{t-2p+1},\ \ldots,\ n_{t-p+1}+n_{t-p-1},\ bn_{t-p}] \qquad (5.22)$$

, then (5.19) becomes

$$\tilde{\underset{\sim}{x}}_t^T\ [1,\ -a_1,\ldots,\ -a_{p-1},\ -\frac{a_p}{b}]^T = \tilde{\underset{\sim}{n}}_t^T\ [1,\ -a_1,\ldots,\ -a_{p-1},\ -a_p]^T$$

, so

$$E[\tilde{\underset{\sim}{x}}_t\tilde{\underset{\sim}{x}}_t^T][1,-a_1,\ldots,-a_{p-1},-\frac{a_p}{b}]^T = E[\tilde{\underset{\sim}{x}}_t\tilde{\underset{\sim}{n}}_t^T][1,-a_1,\ldots,-a_{p-1},-a_p]^T.$$

But from (5.22), we have

$$E[\tilde{\underset{\sim}{x}}_t\tilde{\underset{\sim}{n}}_t^T] = 2\sigma^2\ \mathrm{Diag}(1,\ 1,\ \ldots,\ \frac{b}{2})$$

, that is,

$$\tilde{\underset{\sim}{C}}\ [1,-a_1,\ldots,-a_{p-1},-\frac{a_p}{b}]^T = 2\sigma^2\ [1,-a_1,\ldots,-a_{p-1},-\frac{b}{2}a_p]^T \qquad (5.23)$$

with $\tilde{\underset{\sim}{C}} = E[\tilde{\underset{\sim}{x}}_t\tilde{\underset{\sim}{x}}_t^T]$. Taking $1/b = b/2$, i.e., $b = \pm\sqrt{2}$ implies that $2\sigma^2$ is the minimum eigenvalue of $\tilde{\underset{\sim}{C}}$ and $\tilde{\underset{\sim}{\alpha}}^T = [1,-a_1,\ldots,-a_{p-1},-a_p/b]$ is the corresponding eigenvector. The merits of the estimate based on is that we can achieve a considerable reduction of amounts of computations since the size of $\tilde{\underset{\sim}{C}}$ is $(p+1)\times(p+1)$ while that of $\underset{\sim}{C}$ is $(2p+1)\times(2p+1)$. Another advantage which we found through the experience in computer simulations is that if there are closely spaced frequencies, the estimate of the eigenvector corresponding to the minimum eigenvalue of $\underset{\sim}{C}$ is not necessarily close to $\underset{\sim}{\alpha}$, that is, sometimes the last component $\hat{a}_{2p}$ of the estimate $\hat{\underset{\sim}{\alpha}}$ after normalizing the first component of the eigenvector to 1 becomes 1 whereas the true value is -1. Such a confusion does not occur in the modified method.

Cantoni & Butler (1976) showed that eigenvalues and eigenvectors of any Toeplitz matrix can be classified into two classes. $\tilde{\underset{\sim}{C}}$ in (5.23) corresponds to a submatrix of the decomposition of $\underset{\sim}{C}$. We will discuss this point in detail in a later section.

5.3.3 Statistical Analysis for the Pisarenko Method

In this subsection, we derive the asymptotic variances of the Pisarenko method. Define

$$\begin{aligned} \underset{\sim}{r} &= (r_1, r_2, \ldots, r_m)^T, \quad \underset{\sim}{a} = (a_1, a_2, \ldots, a_m)^T, \\ (\underset{\sim}{R})_{ik} &= r_{i-k} \quad \text{for } i \neq k, \quad (\underset{\sim}{R})_{ii} = r_0 - \sigma^2 \end{aligned} \tag{5.24}$$

with $m = 2p$. Then, (5.21) can be written as

$$\begin{bmatrix} r_0 & \underset{\sim}{r}^T \\ \underset{\sim}{r} & \underset{\sim}{R} + \sigma^2 \underset{\sim}{I}_m \end{bmatrix} \begin{bmatrix} 1 \\ -\underset{\sim}{a} \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 \\ -\underset{\sim}{a} \end{bmatrix}$$

or

$$r_0 - \sigma^2 = \underset{\sim}{r}^T \underset{\sim}{a} \tag{5.25}$$

$$\underset{\sim}{R}\, \underset{\sim}{a} = \underset{\sim}{r} \quad . \tag{5.26}$$

The estimation errors $\Delta \underset{\sim}{a} = \hat{\underset{\sim}{a}} - \underset{\sim}{a}$, $\Delta\sigma^2 = \hat{\sigma}^2 - \sigma^2$ are asymptotically given by

$$\begin{aligned} (\underset{\sim}{R} \Delta \underset{\sim}{a})_k &\approx (\hat{\underset{\sim}{r}} - \hat{\underset{\sim}{R}} \underset{\sim}{a})_k \\ &= \hat{r}_k - \sum_{i=1}^{m} \hat{r}_{k-i} a_i + \hat{\sigma}^2 a_k . \end{aligned} \tag{5.27}$$

Similarly, it follows from (5.25) that

$$\hat{\sigma}^2 \approx \hat{r}_0 - \hat{\underset{\sim}{r}}^T \underset{\sim}{a} - \underset{\sim}{r}^T \Delta \underset{\sim}{a}. \tag{5.28}$$

(See Section 2.2). Combining (5.27), (5.28) and using the periodogram, we obtain

$$\begin{aligned}(\underset{\sim}{R} + \underset{\sim}{a}\underset{\sim}{r}^T)\cdot\Delta\underset{\sim}{a} &\approx \int_{-\pi}^{\pi} B(s)[\underset{\sim}{E}(-s) + \underset{\sim}{a}] I_N(s)\, ds \\ &= \int_{-\pi}^{\pi} B(s)[\underset{\sim}{E}(-s) + \underset{\sim}{a}][I_N(s) - W(s)]\, ds,\end{aligned} \tag{5.29}$$

$$\bar{\sigma}^2 \approx \frac{1}{1 + \underset{\sim}{a}^T \underset{\sim}{a}} \int_{-\pi}^{\pi} B(s)B(-s) I_N(s)\, ds \tag{5.30}$$

where the detailed derivation of (5.30) is deferred to Appendix 5.3. Note the close resemblances of (5.29) and (5.30) to (2.10) and (A-2.1), respectively. Thus the covariance matrix of $\sqrt{N}(\underset{\sim}{R} + \underset{\sim}{a}\underset{\sim}{r}^T)\Delta\underset{\sim}{a}$ is given by

$$N \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} B(s)B(t)[\underset{\sim}{E}(-s)+\underset{\sim}{a}][\underset{\sim}{E}(-t)+\underset{\sim}{a}]\mathrm{Cov}[I_N(s), I_N(t)]\, ds\cdot dt$$

$$\triangleq D = (D_{ik}). \tag{5.31}$$

D_{ik} is obtained by applying (A-5.24) with $F_{ik}(s,t) = B(s)B(t)(e^{jis} + a_i)(e^{jks} + a_k)$. Since $B(\pm\omega_k) = 0$, $k=1,\dots,p$, so that

$$\begin{aligned}\frac{D_{ik}}{\sigma^4} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} [F_{ik}(s,s) + F_{ik}(s,-s)]\, ds \\ &\triangleq g_{ik} + h_{ik}\end{aligned} \tag{5.32}$$

with

$$\begin{aligned}g_{ik} &= a_i a_k + a_i \sum_{n+n'=k} (-a_n)(-a_{n'}) + a_k \sum_{n+n'=i} (-a_n)(-a_{n'}) \\ &\quad + \sum_{n+n'=i+k} (-a_n)(-a_{n'})\end{aligned}$$

$$h_{ik} = a_i a_k \sum_{n-n'=0} (-a_n)(-a_{n'}) + a_i \sum_{n-n'=-k} (-a_n)(-a_{n'})$$

$$+ a_k \sum_{n-n'=-i} (-a_n)(-a_{n'}) + \sum_{n-n'=-i+k} (-a_n)(-a_{n'}). \quad (5.33)$$
$$(n,n'=0,\dots,m)$$

The frequencies ω_k's are estimated by examining peak frequencies of $|B(s)|^{-2}$. Thus the estimation error $\Delta\omega_k = \hat{\omega}_k - \omega_k$ is expressed as

$$\Delta\omega_k \doteq \Delta \underset{\sim}{a}^T \frac{\underset{\sim}{E}(-\omega_k)B'(\omega_k) + \underset{\sim}{E}(\omega_k)B'(-\omega_k)}{2B'(\omega_k)B'(-\omega_k)}, \qquad (5.34)$$

since in (5.10) we put $B(\pm\omega_k) = 0$. Thus the variance is easily obtianed from (5.31) and (5.34). As an example, consider the single sinusoidal case. Then we have $a_1 = 2\cos\omega$, $a_2 = -1$, and

$$\underset{\sim}{R} + \underset{\sim}{a}\underset{\sim}{r}^T = 2|A|^2 \begin{bmatrix} 1+2\cos^2\omega & \cos\omega+2\cos\omega\cdot\cos\omega \\ 0 & 1-\cos2\omega \end{bmatrix}$$
$$g_{11} + h_{11} = 16\cos^4\omega - 12\cos^2\omega + 4 \qquad (5.35)$$
$$g_{12} + h_{12} = g_{21} + h_{21} = 0, \quad g_{22} + h_{22} = 0.$$

Hence, from (5.31) the second component of $(\underset{\sim}{R}+\underset{\sim}{a}\underset{\sim}{r}^T)\Delta\underset{\sim}{a}$ must be zero , i.e., $2|A|^2(1-\cos\omega)\Delta a_2 = 0$. This means that $\Delta a_2 = 0$, or $\hat{a}_2 = a_2 = -1$ *identically* provided that $\omega \neq 0, \pi$. Thus the variance of $\Delta\omega$ in (5.34) is given by

$$E[(\Delta\omega)^2] \doteq \frac{1}{N}\left(\frac{\sigma^2}{2|A|^2}\right)^2 \frac{4\cos^4\omega - 3\cos^2\omega + 1}{\sin^2\omega(1+2\cos^2\omega)^2}. \quad (5.36)$$

The dependence of the variance on N and SNR is same as in the AR method.

In general, we can show that $\hat{a}_{2p} = -1$ and $\hat{a}_k = \hat{a}_{2p-k}$ $(k=1,\dots, p-1)$ *identically*. The proof of this fact is given in Appendix 5.4. This also follows directly from the results of Cantoni & Butler (1976) where it was shown that eigenvectors of a Toeplitz matrix have symmetric or skew symmetric properties. Note that $\underset{\sim}{\alpha}$ in (5.20) is symmetric and so is $\hat{\underset{\sim}{\alpha}}$ as long as $\hat{\underset{\sim}{C}}$ is Toeplitz.

From (5.30) and (A-5.24) it also follows that

$$N.E[(\frac{\Delta\sigma^2}{\sigma^2})^2] \approx \frac{1}{2\pi} \frac{2}{(1 + \underset{\sim}{a}^T\underset{\sim}{a})^2} \int_{-\pi}^{\pi} B^2(s)B^2(-s)\, ds. \quad (5.37)$$

Note that the right hand side of (5.37) does not depend on the signal power. That is, whatever large the power of s_t may be, it influences no effects on the estimation of σ^2.

5.3.4 Comments on the Modified Pisarenko Method

From the result of Appendix 5.4, the Pisarenko estimate also has the symmetric property (5.18) identically. Thus we cannot expect any improvement on performance of the modified estimate. Actually, it can be shown easily but rather tediously that the error expression of the modified estimate corresponding to (5.29) is equal to the one obtained by reducing (5.29) with the symmetric property.

We note from Cantoni & Butler (1976) that $\underset{\sim}{C}$ in (5.21) can be decomposed into

$$\underset{\sim}{C} - \begin{bmatrix} \underset{\sim}{A} & \tilde{\underset{\sim}{r}} & \underset{\sim}{J} \\ \tilde{\underset{\sim}{r}}^T & r_0' & \tilde{\underset{\sim}{r}}^T\underset{\sim}{J} \\ \underset{\sim}{B} & \underset{\sim}{J}\tilde{\underset{\sim}{r}} & \underset{\sim}{J}\underset{\sim}{A}\underset{\sim}{J} \end{bmatrix} \quad (5.38)$$

with

$$\underset{\sim}{A} = \begin{bmatrix} r_0' & \cdots & r_{p-1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ r_{p-1} & \cdots & r_0' \end{bmatrix}, \quad \underset{\sim}{B} - \begin{bmatrix} r_{p+1} & \cdots & r_2 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ r_{2p} & \cdots & r_{p+1} \end{bmatrix}, \quad \underset{\sim}{J} = \begin{bmatrix} & & & 1 \\ & 0 & \cdot^{\cdot^{\cdot}} & \\ & \cdot^{\cdot^{\cdot}} & 0 & \\ 1 & & & \end{bmatrix}$$

and $\tilde{\underline{r}}^T = (r_p,\ldots,r_1)$, $r_0' = r_0 - \sigma^2$. Then, the eigenvalues of $\underline{C}$ belongs to either those of $\underline{A}-\underline{J}\underline{B}$ with skew symmetric eigenvectors or those of

$$\begin{bmatrix} \underline{A}+\underline{J}\underline{B} & \sqrt{2}\,\tilde{\underline{r}} \\ \sqrt{2}\,\tilde{\underline{r}}^T & r_0' \end{bmatrix} = \frac{\tilde{\underline{C}}}{2}$$

with symmetric eigenvectors where $\tilde{\underline{C}}$ is defined by (5.23). Thus we have shown an estimation-theoretic interpretation for Cantoni & Butler's result.

5.3.5 A Special Consideration

As was discussed in Section 5.2.3, the expression (5.36) also becomes invalid as the SNR becomes high with *fixed* N. Although this phenomenon will occur for more general multiple sinusoidal cases, here we perform a special analysis for a single sinusoid.

Using (5.29) with (5.35), we have

$$\Delta a_1 = \frac{(2-4\cos^2\omega)(\hat{r}_1 - r_1) + 2\cos\omega\cdot(\hat{r}_2 - r_2)}{2|A|^2(\ 1 + 2\cos^2\omega\ )}\ .$$

But from (5.34), $(-2\sin\omega)\Delta\omega = \Delta a_1$, so

$$\Delta\omega \approx \frac{\Delta r_1\cos 2\omega - \Delta r_2\cos\omega}{2|A|^2\sin\omega\cdot(1 + 2\cos^2\omega)}\ . \qquad (5.39)$$

Substituting the exact expressions of $\hat{r}_1-r_1$, $\hat{r}_2-r_2$ for $x_t = A\,e^{j\omega t} + A^*e^{-j\omega t} + n_t = 2|A|\sin(\omega t+\phi) + n_t$ into (5.38), we have

$$\sin\omega\cdot(1+2\cos^2\omega)\Delta\omega \approx \frac{1}{N}[\cos 2\omega \sum_{t=1}^{N-1} \cos(\omega(2t+1)+\phi) -$$

$$- \cos\omega \sum_{t=1}^{N-2} \cos(\omega(2t+2)+2\phi) + \cos 2\omega\cdot\cos\omega\ ]$$

$$+ \frac{1}{|A|N} \{n_1[\cos 2\omega \cdot \cos(2\omega+\phi) - \cos\omega \cdot \cos(3\omega+\phi)]$$

$$+ n_2[2\cos(2\omega+\phi)\cos\omega \cdot \cos 2\omega - \cos\omega \cdot \cos(5\omega+\phi)]$$

$$+ n_{N-1}[2\cos(\omega(N-1)+\phi)\cos\omega \cdot \cos 2\omega - \cos\omega \cdot \cos(\omega(N-2)+\phi)]$$

$$+ n_N[\cos 2\omega \cdot \cos(\omega(N-1)+\phi) - \cos\omega \cdot \cos(\omega(N-2)+\phi)]\}$$

$$+ \frac{1}{2|A|^2 N}\left[\cos 2\omega \sum_{t=1}^{N-1} n_t n_{t+1} - \cos\omega \sum_{t=1}^{N-2} n_t n_{t+2}\right] \qquad (5.40)$$

The first term represents a bias. The second term is peculiar to this problem and gives rise to a complication in the expression for the variance. Note that, in general, the variance depends on the phase ϕ. To compare the variance with (5.36), consider a special case $\omega = \pi/2$, $\phi = 0$. Then, (5.40) reduces to

$$\Delta\omega \approx \frac{1}{N|A|}\left(n_1 - n_N \cos\frac{\pi}{2}(N-1)\right) - \frac{1}{2|A|^2 N} \sum_{t=1}^{N-1} n_t n_{t+1}$$

, so that for N even

$$E[(\Delta\omega)^2] \simeq \frac{\sigma^4}{4|A|^4}\frac{1}{N} + \left(\frac{\sigma^2}{|A|^2} - \frac{\sigma^4}{4|A|^4}\right)\frac{1}{N^2} . \qquad (5.41)$$

Of course, the first term of (5.41) coincides with (5.36), but the second term becomes dominant when N is *fixed* and the SNR becomes high. To see the validity of (5.41), numerical simulations were performed. The empirical variances were obtained by averaging the squared errors over 100 different data sets. Fig. 5.6 shows the plots of log $E[(\Delta\omega)^2]$ versus log σ^2 with N = 1000 fixed. The solid line represents the first term of (5.41), the dashed one represents both terms, and o marks show the empirical variances. The agreements are quite good. Fig. 5.7 shows the plots of $\log E[(\Delta\omega)^2]$ versus $\log N$ with $\sigma^2 = 1$ fixed. In this case, the agreements are rather bad for $N < 15$. However, these are inevitable since we have obtained (5.41) under the assumption that Δr_1, Δr_2 are small, i.e., N is sufficiently large.

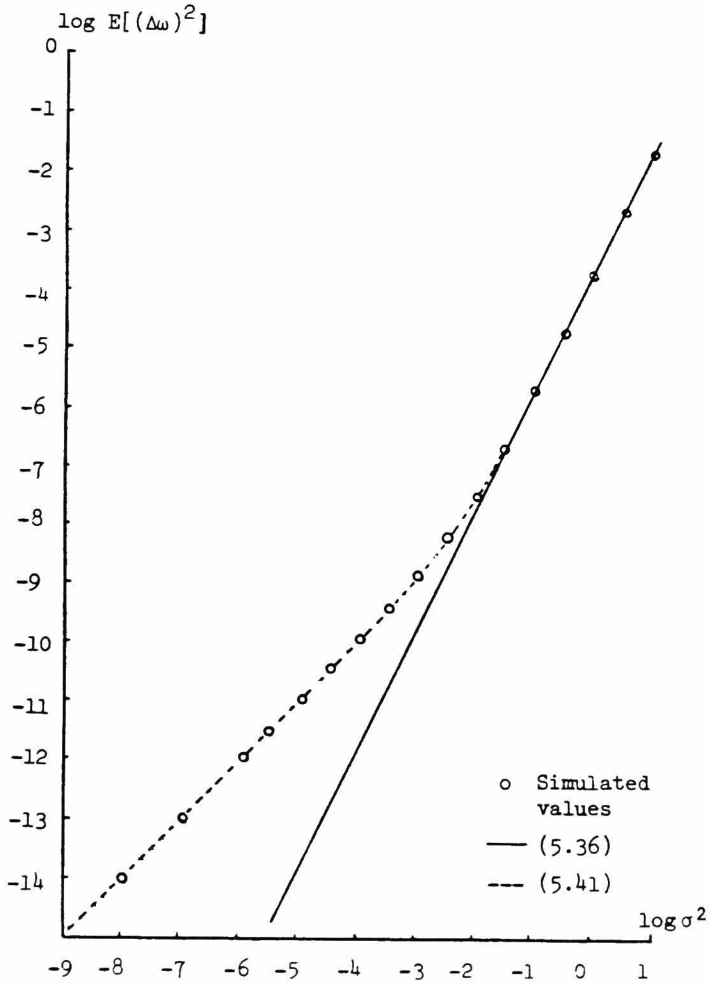

Fig. 5.6 The theoretical and empirical variances of Pisarenko estimate with $\omega = \pi/2$, $\phi = 0$, $N = 1000$.

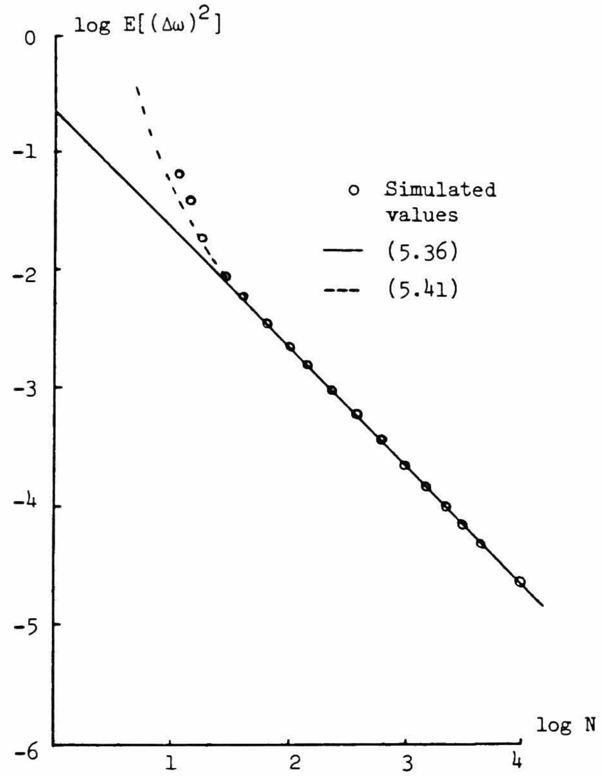

Fig. 5.7 The theoretical and empirical variances of Pisarenko estimate with $\omega = \pi/2$, $\phi = 0$, $\sigma^2 = 1$.

5.4 Discussion and Conclusion

As is noted in Section 5.2.3, the Cramèr-Rao bound for a single sinusoid in white noise is $12/N^3P$. Thus, unless N is small and P is large, the performances of the AR and Pisarenko methods are far from the ideal one. This explains the reason why the performance of the MLS estimator in Chapter 3 becomes disastrously bad as the signal tends to a pure sinusoid. However, this defect does not imply that the two methods are useless. These are easy to compute on-line and provide reasonably good initial estimates for more accurate , but time-consuming maximum likelihood estimates. As for the comparison between the AR and Pisarenko methods, we prefer the AR method because the former has smaller variances than the latter, that is, for $\omega = \pi/4$ and SNR = 10^{-2}, $N \cdot E[(\Delta\omega)^2]$ is about 20 with 15-th order AR method while it is 2500 ! for the Pisarenko method. It is also interesting to note that in the AR method, increasing the order m decreases the variance whereas in the Pisarenko method, this in reverse inflates it. The merit of the Pisarenko method is its high resolvability when there are closely spaced sinusoids (Frost 1977).

This leads us to consider an improved estimate which has the merits of both methods. That is, first fit an AR model to the data and obtain the estimated noise samples $\hat{n}_t$ of n_t. Then form a new data set $\{\tilde{x}_t\}$ such that $\tilde{x}_t = x_t - \lambda\hat{n}_t$ where λ is a suitably selected constant $(0 < \lambda < 1)$, and apply the above procedure to $\{\tilde{x}_t\}$. We expect that $\{\tilde{x}_t\}$ has higher SNR than $\{x_t\}$, so that the variance becomes smaller.

In conclusion, using the periodogram technique, we have understood some aspects of the AR and Pisarenko methods. However, there remains much to investigate about two methods.

Appendix 5.1

In this Appendix, we shall show

$$\int_{-\pi}^{\pi} g(\nu,s)\, E[I_N(s)]\, ds = 1 + O(N^{-1}). \qquad \text{(A-5.1)}$$

Noting $E[n_t] = 0$, it obviously follows that

$$E[I_N(s)] - \frac{1}{2\pi N}\{ \Big|\sum_{t=1}^{N} \sum_{i=1}^{p} (A_i e^{j\omega_i t} + A_i^* e^{-j\omega_i t}) e^{-jts}\Big|^2$$

$$+ E \left| \sum_{t=1}^{N} n_t \, e^{-jts} \right|^2 . \qquad (A\text{-}5.2)$$

As is well known, the second term of (A-5.2) is written as

$$\frac{1}{2\pi} \sum_{k=-N+1}^{N-1} q_k \left(1 - \frac{|k|}{N}\right) e^{-jks} = Q(s) + O(N^{-1}) \qquad (A\text{-}5.3)$$

where $Q(s)$ is the spectrum of $\{n_t\}$ and is defined by $\sum_{k=-\infty}^{\infty} q_k \cdot \exp(-jks)$ $/2\pi$. Introducing the Dirichlet kernel $D_N(s)$ in (4.24), the first term of (A-5.2) becomes

$$\frac{1}{2\pi N} \left| \sum_{i=1}^{p} (A_i D_N(s-\omega_i) + A_i^* D_N(s+\omega_i)) \right|^2 . \qquad (A\text{-}5.4)$$

From the theory of the Fejér kernel, we have

$$\frac{1}{2\pi N} |D_N(s \pm \omega_i)|^2 \rightarrow \delta(s \pm \omega_i) \qquad (A\text{-}5.5)$$

as $N \rightarrow \infty$ where $\delta(.)$ is Dirac's delta function. Also using the fact (4.29) and $\omega_k - \omega_i \neq 2\pi n$ for any i, k and integer n, we obtain

$$\frac{1}{2\pi N} D_N(s - \omega_i) D_N(-s-\omega_k) \rightarrow 0.$$

Thus, apart from the $O(N^{-1})$ term, $E[I_N(s)]$ is given by

$$\sum_{i=1}^{p} |A_i|^2 (\delta(s+\omega_i) + \delta(s-\omega_i)) + Q(s) \triangleq W(s) \qquad (A\text{-}5.6)$$

where $W(s)$ is the spectrum of $\{x_t\}$. Apparently, this result is consistent with (5.2) since

$$\int_{-\pi}^{\pi} W(s) \, e^{jks} \, ds = r_k . \qquad (A\text{-}5.7)$$

Hence, for proving (A-5.1) it is sufficient to show

$$\int_{-\pi}^{\pi} g(\nu, s) W(s) \, ds = 1. \qquad (A\text{-}5.8)$$

Substituting (2.11), (2.35), and (5.5) into the left-hand side of (A-5.8) and noting (A-5.7), we have

$$\sum_{i=0}^{m}\sum_{k=1}^{m}(-a_i)M_k(\nu)r_{k-i} \quad + \quad \sum_{i=0}^{m}\sum_{k=0}^{m}(-a_i)(-a_k)r_{k-i}. \qquad \text{(A-5.9)}$$

The first term is written as $\underset{\sim}{M}^T(\nu)\underset{\sim}{r} + (-\underset{\sim}{a}^T)\underset{\sim}{R}\cdot\underset{\sim}{M}(\nu)$. But this is zero since $\underset{\sim}{M}^T(\nu) = \underset{\sim}{H}^T(\nu)\underset{\sim}{R}^{-1}$ and $\underset{\sim}{R}^{-1}\underset{\sim}{r} = \underset{\sim}{a}$. Similarly, the second term is 1 from (2.3) and (2.13). This completes the proof of (A-5.1).

Appendix 5.2

In this appendix we shall evaluate the asymptotic value of the integral

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} F(s,t)\ \mathrm{Cov}[I_N(s),I_N(t)]\ ds\cdot dt \qquad \text{(A-5.10)}$$

for a sufficiently smooth periodic real function $F(s,t)$. We first seek the expression for $\mathrm{Cov}[I_N(s),I_N(t)]$. For notational simplicity, we put

$$X_1 \triangleq \frac{1}{\sqrt{2\pi N}}\sum_{k=1}^{p}[A_k D_N(s-\omega_k) + A_k^* D_N(s+\omega_k)] \qquad \text{(A-5.11)}$$

$$Y_1 \triangleq \frac{1}{\sqrt{2\pi N}}\sum_{i=1}^{N} n_i\ e^{-jis} \qquad \text{(A-5.12)}$$

and X_2, Y_2 are defined by replacing s by t in (A-5.11), (A-5.12), respectively. Thus, $I_N(s) = |X_1|^2 + 2\ \mathrm{Re}(X_1^*Y_1) + |Y_1|^2$, $I_N(t) = |X_2|^2 + 2\ \mathrm{Re}(X_2^*Y_2) + |Y_2|^2$. From the assumption that $\{n_t\}$ is Gaussian , Y_1 and Y_2 are zero-mean complex Gaussian random variables. Hence, means of their first and third products are zero, i.e., $E[X_1^*Y_1] = E[X_2^*Y_2] = E[(2\cdot\mathrm{Re}(X_1^*Y_1))\cdot|Y_2|^2] = E[(2\cdot\mathrm{Re}(X_2^*Y_2))\cdot|Y_1|^2] = 0$. Consequently, we have

$$\mathrm{Cov}[I_N(s),I_N(t)] = E[(2\cdot\mathrm{Re}(X_1^*Y_1))(2\cdot\mathrm{Re}(X_2^*Y_2))] + \mathrm{Cov}[|Y_1|^2,|Y_2|^2]. \qquad \text{(A-5.13)}$$

In (A-5.13) the first term represents the cross effect due to the signal and the noise and rewritten as

$$X_1^*X_2^* \, E[Y_1Y_2] + X_1X_2 \, E[Y_1^*Y_2^*] + X_1^*X_2 \, E[Y_1Y_2^*] + X_1X_2^* \, E[Y_1^*Y_2]. \qquad \text{(A-5.14)}$$

On the other hand, it follows from (4.23) that

$$E[Y_1Y_2] = \frac{1}{N} Q(s) \, D_N(s+t) + O(N^{-1}) \qquad \text{(A-5.15)}$$

where $D_N(.)$ is defined by (4.24). As is easily seen from the ensuing analysis, the $O(N^{-1})$ term in (A-5.15) does not affect the final result, so that from now on, we discard this term. Thus, (A-5.14) becomes

$$\frac{1}{N} Q(s)[\, X_1^*X_2^* \, D_N(s+t) + X_1X_2 \, D_N(-s-t) + X_1^*X_2 \, D_N(s-t) + X_1X_2^* \, D_N(-s+t)]. \qquad \text{(A-5.16)}$$

The second term of (A-5.13), representing the pure effect due to the noise, is asymptotically equal to

$$\frac{1}{N}[\, |D_N(s+t)|^2 + |D_N(s-t)|^2] \, Q(s)Q(t). \qquad \text{(A-5.17)}$$

(See Section 2.2.) Using the approximation (A-5.5) in the integral (A-5.10), the contribution due to (A-5.17) is given by

$$\frac{2\pi}{N} \int_{-\pi}^{\pi} [F(s,s) + F(s,-s)]Q^2(s) \, ds. \qquad \text{(A-5.18)}$$

Next we examine the contribution due to (A-5.16). It is sufficient to consider the first and third term of (A-5.16). Upon substituting (A-5.11) into these terms and performing the integration for (A-5.10), we encounter the following type of integrals

$$\frac{1}{2\pi N^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} D_N(\mu-s) D_N(\nu\mp t) D_N(s\pm t) F(s,t) Q(s)\, ds\cdot dt. \quad \text{(A-5.19)}$$

From the assumption that $F(s,t)$ is periodic $F(s,t)Q(s)$ can be expressed as the following Fourier series:

$$F(s,t)Q(s) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \sum_{n'=-\infty}^{\infty} f_{n,n'}\, e^{-j(sn+tn')}. \quad \text{(A-5.20)}$$

Substituting (A-5.20) and (4.24) into (A-5.19), we have

$$\frac{1}{(2\pi N)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \sum_{k=1}^{N} e^{-j(\mu-s)k} \sum_{k'=1}^{N} e^{-j(\nu\mp t)k'}$$

$$\times \sum_{k''=1}^{N} e^{-j(s\pm t)k''} \sum_{n,n'} f_{n,n'}\, e^{-j(sn+tn')}\, ds\cdot dt$$

$$= \frac{1}{(2\pi N)^2} \sum_{k,k',k'',n,n'} f_{n,n'}\, e^{-j(\mu k+\nu k')}$$

$$\times \int_{-\pi}^{\pi} e^{-j(-k+k''+n)s}\, ds\cdot \int_{-\pi}^{\pi} e^{-j(\mp k'\pm k''+n')t}\, dt$$

$$= \frac{1}{N^2} \sum_{k,k',k'',n,n'} f_{n,n'}\, e^{-j(\mu k+\nu k')}\, \delta_{-k+k''+n,0}$$

$$\times\, \delta_{\mp k'\pm k''+n',0}\, .$$

$$= \frac{1}{N^2} \sum_{n,n'} f_{n,n'}\, e^{-j(\mu n\pm\nu n')} \sum_{k''\varepsilon E_{n,n'}} e^{-j(\mu+\nu)k''} \quad \text{(A-5.21)}$$

$$= \frac{2\pi}{N} F(\mu,\pm\nu) Q(\mu) D_N(\mu+\nu) + O(N^{-2}). \quad \text{(A-5.22)}$$

In (A-5.24) $E_{n,n'}$ denotes the common set of [1-n,N-n], [1 n',N n'], and [1,N] and differs from [1,N]. This difference causes an extra term in (A-5.22). But this term is of order N^{-2} because of the smoothness of $F(s,t)Q(s)$.

Consequently, it follows by (4.29) that the asymptotic value of (A-5.19) is

$$\begin{cases} \frac{2\pi}{N} F(\mu,\mp\mu)Q(\mu) & \text{if } \mu+\nu = 0 \\ O(N^{-2}) & \text{otherwise .} \end{cases} \qquad \text{(A-5.23)}$$

Applying this result to calculate the contribution of the first and third term of (A-5.16), we obtain

$$\frac{2\pi}{N} \sum_{k=1}^{p} |A_k|^2 \{ F(\omega_k,-\omega_k) + F(-\omega_k,\omega_k)\} Q(\omega_k)$$

corresponding to the first term, and

$$\frac{2\pi}{N} \sum_{k=1}^{p} |A_k|^2 \{ F(\omega_k,\omega_k) + F(-\omega_k,-\omega_k)\} Q(_k)$$

corresponding to the third term, respectively. Hence, the asymptotic value of (A-5.10) is

$$\frac{2\pi}{N} \{ \sum_{k=1}^{p} 2|A_k|^2 [F(\omega_k,-\omega_k) + F(-\omega_k,\omega_k) + F(\omega_k,\omega_k) + F(-\omega_k,-\omega_k)]Q(\omega_k)$$
$$+ \int_{-\pi}^{\pi} [F(s,s) + F(s,-s)]Q^2(s)\, ds \}. \qquad \text{(A-5.24)}$$

Using (A-5.6), (A-5.24) is more compactly written as

$$\frac{2\pi}{N} \int_{-\pi}^{\pi} [F(s,s) + F(s,-s)]Q(s)[2W(s) - Q(s)]\, ds. \quad \text{(A-5.25)}$$

Appendix 5.3

Derivation of (5.30)

We first show that $\underset{\sim}{R} + \underset{\sim\sim}{ar}^T$ is invertible. From (5.26) $\underset{\sim}{R} + \underset{\sim\sim}{ar}^T$ $= (\underset{\sim}{I} + \underset{\sim}{a}\cdot\underset{\sim}{a}^T)\underset{\sim}{R}$, but $|\underset{\sim}{I} + \underset{\sim}{a}\cdot\underset{\sim}{a}^T| = 1 + a_1^2 + \ldots + a_m^2 > 0$, so $|\underset{\sim}{R} + \underset{\sim\sim}{ar}^T|$ > 0, that is, $(R + ar^T)^{-1}$ exists.

Substitution of (5.29) into (5.28) gives

$$\hat{\sigma}^2 = \int_{-\pi}^{\pi} B(s)[1 - \underset{\sim}{r}^T(\underset{\sim}{R} + \underset{\sim\sim}{ar}^T)^{-1}(\underset{\sim}{E}(-s) + \underset{\sim}{a})]I_N(s)\,ds.$$

From the well known matrix inversion lemma; $(\underset{\sim}{A}+\underset{\sim\sim}{BC})^{-1} = \underset{\sim}{A}^{-1}-\underset{\sim}{A}^{-1}\underset{\sim}{B}(\underset{\sim}{I}$ $+\underset{\sim\sim}{CA}^{-1}\underset{\sim}{B})^{-1}\underset{\sim\sim}{CA}$, we have

$$(\underset{\sim}{I} + \underset{\sim}{a}\cdot\underset{\sim}{a}^T)^{-1} = \underset{\sim}{I} - \frac{1}{1 + \underset{\sim}{a}^T\underset{\sim}{a}}\,\underset{\sim}{a}\cdot\underset{\sim}{a}^T. \qquad \text{(A-5.26)}$$

Thus,

$$\underset{\sim}{r}^T(\underset{\sim}{R} + \underset{\sim\sim}{ar}^T)^{-1} = \underset{\sim}{r}^T\underset{\sim}{R}^{-1}(\underset{\sim}{I} + \underset{\sim}{a}\cdot\underset{\sim}{a}^T)^{-1}$$
$$= \underset{\sim}{a}^T(I - \frac{\underset{\sim}{a}\cdot\underset{\sim}{a}^T}{1 + \underset{\sim}{a}^T\underset{\sim}{a}}) = \frac{\underset{\sim}{a}^T}{1 + \underset{\sim}{a}^T\underset{\sim}{a}}.$$

But from $\underset{\sim}{a}^T\underset{\sim}{E}(s) = 1 - B(s)$, we have

$$1 - \underset{\sim}{r}^T(\underset{\sim}{R} + \underset{\sim\sim}{ar}^T)^{-1}(\underset{\sim}{E}(-s) + \underset{\sim}{a}) = \frac{B(-s)}{1 + \underset{\sim}{a}^T\underset{\sim}{a}}.$$

Hence, (5.30) follows.

Appendix 5.4

In this Appendix, we want to show $\hat{a}_{2p} = -1$ and $\hat{a}_k = \hat{a}_{2p-k}$ (k=1, ...,p-1). Define the left hand side of (5.29) as $\underset{\sim}{b} = (b_1, b_2, \ldots, b_{2p})$. From the symmetric property (5.18) we have $B(-s)e^{-j2ps} = B(s)$. Noting $I_N(-s) = I_N(s)$, it follows that $b_k = b_{2p-k}$ (k=1,...,p-1) and $b_{2p} = 0$. That is,

$$\underset{\sim}{b} = (b_1,b_2,\ldots,b_{p-1},b_p,b_{p+1},\ldots,b_1,0)^T.$$

From (A-5.26) we have

$$\underset{\sim}{R}\cdot\Delta\underset{\sim}{a} = (\underset{\sim}{I} + \underset{\sim}{a}\cdot\underset{\sim}{a}^T)^{-1}\,\underset{\sim}{b} = \underset{\sim}{b} - \frac{\underset{\sim}{a}^T\underset{\sim}{b}}{1 + \underset{\sim}{a}^T\underset{\sim}{a}}\,\underset{\sim}{a}$$
$$\triangleq \underset{\sim}{d} \qquad \text{(A-5.27)}$$

Put

$$\underset{\sim}{d} = \begin{bmatrix} \bar{\underset{\sim}{d}} \\ \cdots\cdots \\ d_{2p} \end{bmatrix} \qquad \underset{\sim}{J} = \underbrace{\begin{bmatrix} & & 1 \\ 0 & \cdot^{\cdot^{\cdot}} & 0 \\ 1 & & \end{bmatrix}}_{2p-1} \Big\} 2p-1 \tag{A-5.28}$$

, then

$$\underset{\sim}{J}\cdot\bar{\underset{\sim}{d}} = \bar{\underset{\sim}{d}} \tag{A-5.29}$$

$$d_{2p} = \frac{\underset{\sim}{a}^T\underset{\sim}{b}}{1 + \underset{\sim}{a}^T\underset{\sim}{a}} = \underset{\sim}{r}^T\cdot\Delta\underset{\sim}{a} \tag{A-5.30}$$

where we use the facts $b_{2p} = 0$ and $\underset{\sim}{a}^T(\underset{\sim}{R} + \underset{\sim}{a}\underset{\sim}{r}^T) = (1 + \underset{\sim}{a}^T\underset{\sim}{a})\underset{\sim}{r}$. Put

$$\underset{\sim}{R} = \begin{bmatrix} \bar{\underset{\sim}{R}} & \vdots & \bar{\underset{\sim}{r}} \\ \cdots & \cdots & \cdots \\ \bar{\underset{\sim}{r}}^T & \vdots & r_0' \end{bmatrix} , \quad \Delta\underset{\sim}{a} = \begin{bmatrix} \Delta\bar{\underset{\sim}{a}} \\ \cdots\cdots \\ \Delta a_{2p} \end{bmatrix} \tag{A-5.31}$$

with $r_0' = r_0 \quad \sigma^2$, then (A-5.27) is written as

$$\begin{bmatrix} \bar{\underset{\sim}{R}} & \vdots & \bar{\underset{\sim}{r}} \\ \cdots & \cdots & \cdots \\ \bar{\underset{\sim}{r}}^T & \vdots & r_0' \end{bmatrix} \begin{bmatrix} \Delta\bar{\underset{\sim}{a}} \\ \cdots\cdots \\ \Delta a_{2p} \end{bmatrix} = \begin{bmatrix} \bar{\underset{\sim}{d}} \\ \cdots\cdots \\ d_{2p} \end{bmatrix} \tag{A-5.32}$$

That is,

$$\bar{\underset{\sim}{R}}\cdot\Delta\bar{\underset{\sim}{a}} + \bar{\underset{\sim}{r}}\cdot\Delta a_{2p} = \bar{\underset{\sim}{d}} \tag{A-5.32}$$

$$\bar{\underset{\sim}{r}}^T\Delta\bar{\underset{\sim}{a}} + r_0'\, a_{2p} = d_{2p}\,. \tag{A-5.33}$$

On the other hand, since from (A-5.31) $\bar{\underset{\sim}{r}} = (r_{2p-1} \cdots r_1)^T$, (A-5.30) becomes

$$d_{2p} = (\underset{\sim}{J}\cdot\bar{\underset{\sim}{r}})^T\Delta\bar{\underset{\sim}{a}} + r_{2p}\Delta a_{2p}. \tag{A-5.34}$$

From (A-5.33), (A-5.34) we obtain

$$\bar{\underset{\sim}{r}}^T(\,\Delta\bar{\underset{\sim}{a}} - \underset{\sim}{J}\,\Delta\bar{\underset{\sim}{a}}\,) = (r_{2p} - r_0')\, a_{2p}. \tag{A-5.35}$$

Now we consider two cases separately.

i) First assume $r_{2p} - r_0' \neq 0$. Substituting (A-5.35) into (A-5.32), we have

$$\bar{R}\cdot\Delta\bar{a} + \frac{1}{r_{2p} - r'_0}\,\bar{r}\cdot\bar{r}^T(\Delta\bar{a} - J\cdot\Delta\bar{a}) = \bar{d} \qquad \text{(A-5.36)}$$

Multiplying J from the left in (A-5.36) and noting that $J\cdot\bar{R} = \bar{R}\cdot J$ and $J\cdot\bar{d} = \bar{d}$, it follows that

$$\bar{R}\cdot J\cdot\Delta\bar{a} + J\cdot r\cdot r^T\,\frac{\Delta\bar{a} - J\Delta\bar{a}}{r_{2p} - r'_0} = \bar{d}. \qquad \text{(A-5.37)}$$

Subtracting (A-5.36) from (A-5.37) we have

$$\left(\bar{R} + \frac{\bar{r}\cdot\bar{r}^T - J\cdot\bar{r}\cdot\bar{r}^T}{r_{2p} - r'_0}\right)(\Delta\bar{a} - J\cdot\Delta\bar{a}) = 0. \qquad \text{(A-5.38)}$$

We write $R\cdot a = r$ as

$$\begin{bmatrix} \bar{R} & \bar{r} \\ \bar{r}^T & r'_0 \end{bmatrix} \begin{bmatrix} \bar{a} \\ -1 \end{bmatrix} = \begin{bmatrix} J\cdot\bar{r} \\ r_{2p} \end{bmatrix}$$

or

$$\bar{R}\cdot\bar{a} - \bar{r} = J\cdot\bar{r} \qquad \text{(A-5.39)}$$

$$\bar{r}^T\bar{a} - r'_0 = r_{2p} \quad . \qquad \text{(A-5.40)}$$

Hence from (A-5.39) we have

$$\bar{R} + \frac{\bar{r} - J\cdot\bar{r}}{r_{2p} - r'_0}\,\bar{r}^T = [I_{2p-1} + \frac{\bar{r} - J\cdot\bar{r}}{r_{2p} - r'_0}(\bar{a} - \bar{R}^{-1}J\cdot\bar{r})^T]\bar{R}. \qquad \text{(A-5.41)}$$

Define

$$\bar{R}^{-1}J\cdot\bar{r} = v = (v_1,\dots,v_{2p-1})^T \qquad \text{(A-5.42)}$$

, then it easily follows that

$$\left| I_{2p-1} + \frac{\bar{r} - J\cdot\bar{r}}{r_{2p} - r'_0}(\bar{a} - v)^T \right| = 1 + \frac{1}{r_{2p} - r'_0}(\bar{r} - J\cdot\bar{r})(\bar{a} - v)$$

$$= \frac{1}{r'_0 - r_{2p}}[\, r'_0 \sum_{i=1}^{2p-1} r_i v_i - (r_{2p} - \sum_{i=1}^{2p-1} r_{2p-i} v_i)], \qquad \text{(A-5.43)}$$

since $J\cdot\bar{a} = \bar{a}$. Note that (A-5.42) implies that v is a solution of the (2p-1)-th order normal equation for $\{s_t\}$, so that the (2p-1)-th order partial autocorrelation coefficient

$$\rho_{2p-1} = \frac{r_{2p} - \sum_{i=1}^{2p-1} r_{2p-i}v_i}{r'_0 - \sum_{i=1}^{2p-1} r_i v_i} \tag{A-5.44}$$

satisfies $|\rho_{2p-1}| < 1$ (Burg 1975). This means that the right hand side of (A-5.43) is nonzero so that the matrix in (A-5.38) is invertible. That is, $\Delta\bar{\underset{\sim}{a}} = \underset{\sim}{J}\cdot\Delta\bar{\underset{\sim}{a}}$ and $\Delta a_{2p} = 0$ or $\hat{a}_k = \hat{a}_{2p-k}$ (k=1,..., p-1) and $\hat{a}_{2p} = -1$.

ii) Assume $r_{2p} - r'_0 = 0$. Since $r_{2p} = 2\sum_{i=1}^{p} |A_i|^2 \cos(2p\omega_i)$ and $r'_0 = 2\sum_{i=1}^{p}|A_i|^2$, we must have $2p\omega_i = 2\pi q_i$ where q_i (i=1,...,p) are integers. Thus we have

$$\bar{\underset{\sim}{r}} = \underset{\sim}{J}\cdot\bar{\underset{\sim}{r}} . \tag{A-5.45}$$

Using (A-5.32) and (A-5.45), we have

$$\bar{\underset{\sim}{R}}(\Delta\bar{\underset{\sim}{a}} - \underset{\sim}{J}\Delta\bar{\underset{\sim}{a}}) = 0.$$

Hence $\Delta\bar{\underset{\sim}{a}} = \underset{\sim}{J}\Delta\bar{\underset{\sim}{a}}$ follows. Eliminating $\Delta\bar{\underset{\sim}{a}}$ in (A-5.32) and (A-5.33) we obtain

$$(\bar{\underset{\sim}{r}}^T\underset{\sim}{v} - r'_0)\, \Delta a_{2p} = \underset{\sim}{v}^T\bar{\underset{\sim}{d}} - d_{2p} \tag{A-5.46}$$

where $r'_0 - \bar{\underset{\sim}{r}}^T\underset{\sim}{v} = r'_0 - (\underset{\sim}{J}\cdot\bar{\underset{\sim}{r}})^T\underset{\sim}{v}$ is the (2p-1)-th order mean squared prediction error variance for $\{s_t\}$ and is positive. On the other hand, from (A-5.39) we have $\bar{\underset{\sim}{a}} = 2\underset{\sim}{v}$. Hence, $d_{2p} - \underset{\sim}{v}^T\bar{\underset{\sim}{d}} = d_{2p} - \bar{\underset{\sim}{a}}^T\bar{\underset{\sim}{d}}/2 = d_{2p} - (\underset{\sim}{a}^T\underset{\sim}{d} - a_{2p}d_{2p})/2 = (d_{2p} - \underset{\sim}{a}^T\underset{\sim}{d})/2$. But from (A-5.27) and (A-5.30), $\underset{\sim}{a}^T\underset{\sim}{d} = \underset{\sim}{r}^T\Delta\underset{\sim}{a} = d_{2p}$. Thus from (A-5.46) we obtain $\Delta a_{2p} = 0$.

Chapter 6

Some Problems on Multivariate AR Processes

6.1 Introduction

In this chapter, we consider some problems on parameter estimation of multivariate AR processes. First, statistical properties of the multivariate AR spectral estimator are derived. The result is a generalization of that of Akaike (1969 b). Second, we turn our attention to partial autocorrelation matrices and their normalized forms due to Morf *et al.* (1978 a,b). The latter ones were introduced from algorithmic considerations, but here we show that they have a desirable statistical property which is a multivariate version of that of Quenouille (1947). Third, we obtain the order distributions determined by AIC. This generalizes the result of Shibata (1976). It is interesting that the probability of selecting the correct order increases to 1 as the number of variates increases. This fact is rather contrary to our intuition. Fourth, we derive practical computational algorithms of a new estimation method of Pagano (1978) based on the relations between periodic and multivariate autoregressions. Especially, we demonstrate a circular lattice structure of the algorithm which is a direct generalization of the lattice structure of Itakura & Saito (1971). The circular lattice filtering may be a powerful technique for processing multivariate time series since it can simultaneously perform whitening and orthogonalization of the time series. Unlike most of the conventional multivariate processing methods, this new technique contains no matrix manipulations. Lastly

, we suggest an application to multichannel data compression.

6.2 Multivariate AR Spectral Estimation

6.2.1 Introduction

In recent years, multivariate AR spectral analysis method has been successfully applied to many actual problems (Akaike & Nakagawa 1972). But as for the statistical properties of this method, as far as the author is aware, there are little substantial works except for a comment of Parzen (1970). The study of this method is just a multivariate version of that of Akaike (1969 b), but for this purpose we must fully know the statistical properties of the parameter estimates of multivariate AR processes. The well known paper of Mann & Wald (1943) discussed these problems where they derived the asymptotic distributions of the estimates of the coefficient matrices of an AR process, but did not mention any of the statistical properties of the estimate of a residual covariance matrix, another important quantity in AR processes.

In this section, using the periodogram technique, we first derive several properties of the parameter estimates. Then using these , we generalize the work of Akaike (1969 b). To see the validities of the obtained results, we perform numerical simulations and compare the simulational results and the theoretical ones. We also see that if the fitted order is taken large as compared with the true one, the resulting statistical properties of the multivariate AR spectral estimate resemble with those of the multivariate Blackman-Tukey method.

6.2.2 Multivariate AR Processes

Let us consider the following d-variates m-th order Gaussian AR process

$$\underset{\sim}{X}_t - A_1\underset{\sim}{X}_{t-1} - \dots - A_m\underset{\sim}{X}_{t-m} = \underset{\sim}{U}_t \tag{6.1}$$

where $A_1, A_2, \dots, A_m$ are d×d AR coefficient matrices, $\{\underset{\sim}{U}_t\}$ is a sequence of Gaussian white noise d-vectors with mean $\underset{\sim}{0}$ and covariance matrix W which from now on we call "residual covariance". That is, $E[\underset{\sim}{U}_t] = \underset{\sim}{0}$, $E[\underset{\sim}{U}_t\underset{\sim}{U}_s^T] = W\ \delta_{t,s}$.(W is positive definite.)

Corresponding to (2.3) for uni-variate case, $A_1, \dots, A_m$ satisfies the following Yule-Walker equations

$$[A_1, A_2, \dots, A_m]\begin{bmatrix} R_0 & R_1 & \cdots & R_{m-1} \\ R_{-1} & R_0 & \cdots & R_{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ R_{-m+1} & R_{-m+2} & \cdots & R_0 \end{bmatrix} = [R_1, R_2, \dots, R_m] \tag{6.2}$$

where we define $R_k = E[\underset{\sim}{X}_{t+k}\underset{\sim}{X}_t^T]$. We denote the dm×dm grand matrix in the left hand side of (6.2) by $\underset{\sim}{R}$ whose (i,k)-th block matrix $\underset{\sim}{R}_{ik}$ is R_{k-i}. Similarly, the (i,k)-th block matrix $\underset{\sim}{P} = \underset{\sim}{R}^{-1}$ is denoted by $\underset{\sim}{P}_{ik}$. The residual covariance satisfies

$$W = R_0 - A_1R_{-1} - \dots - A_mR_{-m}. \tag{6.3}$$

The spectral density matrix F(s) of (6.1) is given by

$$F(s) = \frac{1}{2\pi} A^{-1}(s)W\cdot A^{-T}(-s) \tag{6.4}$$

where A^{-T} means $(A^{-1})^T$ and A(s) is defined by

$$A(s) = I - \sum_{k=1}^{m} A_k\, e^{-jks}.\text{(I : d×d unit matrix)} \tag{6.5}$$

To ensure the stationarity of (6.1), we assume that

$$|I - \sum_{k=1}^{m} A_k z^{-k}| \neq 0 \quad \text{for} \quad |z| \leqq 1. \tag{6.6}$$

When a set of data $\{\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_N\}$ is available, we first estimate R_k by the following consistent estimate

$$\hat{R}_k = \frac{1}{N} \sum_{t=1}^{N-k} \underset{\sim}{X}_{t+k} \underset{\sim}{X}_t^T, \quad \hat{R}_{-k} = \hat{R}_k^T \quad (k \geqq 0). \tag{6.7}$$

Defining the periodogram matrix by

$$I_N(s) = \frac{1}{2\pi N} [\sum_{t=1}^{N} \underset{\sim}{X}_t e^{-jts}][\sum_{t=1}^{N} \underset{\sim}{X}_t^T e^{jts}]. \tag{6.8}$$

Corresponding to (2.9) we have

$$R_k = \int_{-\pi}^{\pi} F(s) e^{jks} ds, \quad \hat{R}_k = \int_{-\pi}^{\pi} I_N(s) e^{jks} ds. \tag{6.9}$$

6.2.3 The Error Covariances of the Estimated AR Coefficient Matrices

Substituting $\hat{R}_k$ into R_k in (6.2) gives the estimate $\hat{A}_i$ of A_i. By the same method as in Chapter 2, the estimation error $\Delta A_i = \hat{A}_i$ A_i $(i=1,\ldots,m)$ are asymptotically expressed as

$$[\Delta A_1, \ldots, \Delta A_m] \underset{\sim}{R} \approx [\hat{R}_1 - \sum_{k=1}^{m} A_k \hat{R}_{1-k}, \ldots, \hat{R}_m - \sum_{k=1}^{m} A_k \hat{R}_{m-k}].$$

From (6.5), (6.9) we have

$$\Delta A_i \approx \int_{-\pi}^{\pi} A(s) I_N(s) \sum_{k=1}^{m} \underset{\sim}{P}_{ki} e^{jks} ds. \tag{6.10}$$

We denote the (i,k)-th element of $A(s)$ by $(A(s))_{ik} = a_{ik}(s)$. We rewrite (6.8) as

$$(I_N(s))_{ik} = \frac{1}{2\pi N} J_i(s)J_k(-s), \quad J_k(s) = \left(\sum_{t=1}^{N} \underset{\sim}{X}_t e^{-jts} \right)_k. \quad (6.11)$$

Then, the error covariance betwenn $(\hat{A}_{m_1})_{i_1j_1}$ and $(\hat{A}_{m_2})_{i_2j_2}$ is given by

$$E[(\Delta A_{m_1})_{i_1j_1}(\Delta A_{m_2})_{i_2j_2}] \doteq \frac{1}{(2\pi N)^2} E\{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \sum_{k} \sum_{n,n'} a_{i_1n}(s)$$

$$\times J_n(s)J_{n'}(-s)(\underset{\sim}{P}_{km_1})_{n'j_1} \sum_{k'} \sum_{p,p'} a_{i_2p}(t)J_p(t)J_{p'}(-t)$$

$$\times (P_{k'm_2})_{p'j_2} e^{jk't} \} ds\cdot dt \qquad (6.12)$$

$$(1 \leqq k,k' \leqq m \ , \ 1 \leqq n,n',p,p' \leqq d)$$

Similar to (4.23), it is well known (Brillinger 1975, page 73) that

$$E[J_p(s)J_q(t)] = 2\pi f_{pq}(s)D_N(s+t) + O(1) \qquad (6.13)$$

with $f_{pq}(s) = (F(s))_{pq}$. The second term of (6.13) is uniform in s, t so that its contributions to the integrals below can be neglected.

From the assumption $J_k(s)$ is Gaussian, so that from the property of the fourth moment of Gaussian random variables and (6.13), we obtain

$$E[J_n(s)J_{n'}(-s)J_p(t)J_{p'}(-t)] \doteq (2\pi N)^2 f_{nn'}(s)f_{pp'}(t)$$

$$+ (2\pi)^2 f_{np}(s)f_{n'p'}(-s)|D_N(s+t)|^2 + (2\pi)^2 f_{np'}(s)f_{n'p}(-s)|D_N(s-t)|^2 \qquad (6.14)$$

The contribution of the first term to (6.12) is zero by (6.2). Using (A-5.5), the contribution of the second term is

$$\frac{2\pi}{N} \sum_{k,n,n'} \sum_{k',p,p'} \int_{-\pi}^{\pi} a_{i_1n}(s)f_{np}(s)(\underset{\sim}{P}_{km_1})_{n'j_1} a_{i_2p}(-s)f_{n'p'}(-s)$$

$$\times\ (\underset{\sim}{P}_{k'm_2})_{p'j_2}\ e^{j(k-k')s}\ ds. \tag{6.15}$$

But

$$\sum_{n,p} a_{i_1 n}(s) f_{np}(s) a_{i_2 p}(-s) = (A(s)F(s)A^T(-s))_{i_1 i_2}$$
$$= (2\pi)^{-1}(W)_{i_1 i_2}. \tag{6.16}$$

Noting $F(s) = F^T(-s)$, $\underset{\sim}{R}^T_{kk'} = \underset{\sim}{R}_{k'k}$, the remaining part of (6.15) becomes

$$\sum_{k,k',n',p'} (\underset{\sim}{P}_{km_1})_{n'j_1} (\underset{\sim}{P}_{k'm_2})_{p'j_2} (\underset{\sim}{R}_{k'k})_{p'n'}$$

$$= \sum_{k',p'} \Big(\sum_k \underset{\sim}{R}_{k'k} \cdot \underset{\sim}{R}_{km_1} \Big)_{p'j_1} (\underset{\sim}{P}_{k'm_2})_{p'j_2}$$

$$= \Big(\sum_k \underset{\sim}{I}_{k'm_1} \cdot \underset{\sim}{P}_{k'm_2} \Big)_{j_1 j_2} = (\underset{\sim}{P}_{m_1 m_2})_{j_1 j_2} \tag{6.17}$$

where $\underset{\sim}{I}_{ik}$ is the (i,k)-th block matrix of the $dm\times dm$ unit matrix $\underset{\sim}{I}$ and we used the relation $\underset{\sim}{R}\cdot\underset{\sim}{P} = \underset{\sim}{I}$. By the same argument , we have the contribution of the third term as

$$\frac{2\pi}{N} \sum_{k,k} \int_{-\pi}^{\pi} (A(s)F(s)\underset{\sim}{P}_{k'm_2})_{i_1 j_2} (A(s)F(s)\underset{\sim}{P}_{km_1})_{i_2 j_1} e^{j(k+k')s}\ ds. \tag{6.18}$$

Rewriting the integral (6.18) into a complex integral by the transformation $z = \exp(js)$ and using Cauchy's theorem, we can easily see that this integral is zero, since from (6.4), (6.5), and (6.6), $2\pi A(s)F(s) = W\cdot A^{-T}(-s) = W\cdot \mathrm{Adj}(I-A_1z-\dots-A_mz^m)^T/|I-A_1z-\dots-A_mz^m|$ is regular inside the unit circle and $k+k'-1 > 0$. Thus we have the desired result

$$N\cdot E[(\Delta A_{m_1})_{i_1 j_1}(\Delta A_{m_2})_{i_2 j_2}] \approx (W)_{i_1 i_2}(\underset{\sim}{P}_{m_1 m_2})_{j_1 j_2}. \tag{6.19}$$

6.2.4 The Error Covariances of the Estimated Residual Matrix

In this subsection, we derive new results about the statistical properties of the estimated residual matrix. Since the estimate takes the form of

$$\hat{W} = \hat{R}_0 - \sum_{k=1}^{m} \hat{A}_k \hat{R}_{-k} \tag{6.20}$$

, the estimation error is asymptotically expressed as

$$\Delta W \approx \Delta R_0 - \sum_{i=1}^{m} (\Delta A_i R_{-i} + A_i \Delta R_{-i})$$

$$= \int_{-\pi}^{\pi} A(s) I_N(s) ds - \sum_{i=1}^{m} \Delta A_i R_{-i} - W. \tag{6.21}$$

Hence, we have

$$E[(\Delta A_{m_1})_{i_1 j_1} (\Delta W)_{i_2 j_2}] \sim \int_{-\pi}^{\pi} E[(\Delta A_{m_1})_{i_1 j_1} (A(s) I_N(s))_{i_2 j_2}]\, ds$$

$$- \sum_{i=1}^{m} E[(\Delta A_{m_1})_{i_1 j_1} (\Delta A_i R_{-i})_{i_2 j_2}] - E[(\Delta A_{m_1})_{i_1 j_1}](W)_{i_2 j_2}. \tag{6.22}$$

Substituting (6.10) into the first term of (6.22) and using the same technique as in the calculation of (6.21), the first term is asymptotically equal to

$$\frac{1}{N}(W)_{i_1 i_2} (A_{m_1})_{j_2 j_1} + E[(\Delta A_{m_1})_{i_1 j_1}](W)_{i_2 j_2}. \tag{6.23}$$

From (6.19), the second term becomes

$$\sum_{i,p} E[(\Delta A_{m_1})_{i_1 j_1} (\Delta A_i)_{i_2 p}](R_{-i})_{p j_2}$$

$$= \frac{1}{N}(W)_{i_1 i_2} \Big(\sum_i \underset{\sim}{P}_{m_1 i} \cdot R_{-i} \Big)_{j_1 j_2}$$

$$= \frac{1}{N}(W)_{i_1 i_2} (A_{m_1}^T)_{j_1 j_2} \tag{6.24}$$

where we used the relations $R^T_{-i} = R_i$, $P^T_{ik} = P_{ki}$, and

$$\sum_i P_{ki} R_{-i} = A^T_k. \tag{6.25}$$

Hence, we obatin

$$N \cdot E[(\Delta A_{m_1})_{i_1 j_1} (\Delta W)_{i_2 j_2}] = 0. \tag{6.26}$$

This means that the estimation errors of the AR coefficient matrices and the estimation error of the residual matrix are asymptotically uncorrelated. This is a natural generalization of the corresponding result (2.28) for uni-variate case.

Substituting (6.10) into ΔA_i in (6.21) and using (6.25), ΔW is expressed as

$$\Delta W = \int_{-\pi}^{\pi} A(s) I_N(s) A^T(-s)\, ds - W. \tag{6.27}$$

This is a multivariate version of (A-2.1). Thus,

$$\begin{aligned} E[(\Delta W)_{i_1 j_1} (\Delta W)_{i_2 j_2}] &\approx \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} E[(A(s) I_N(s) A^T(-s))_{i_1 j_1} (A(t) I_N(t) \\ &\times A^T(-t))_{i_2 j_2}]\, ds \cdot dt - E[(\hat{W})_{i_1 j_1}](W)_{i_2 j_2} - (W)_{i_1 j_1} \\ &\times E[(\hat{W})_{i_2 j_2}] + (W)_{i_1 j_1} (W)_{i_2 j_2}. \end{aligned} \tag{6.28}$$

Writing the first term of (6.28) componentwise, using (6.14), and performing the same calculation as in (6.12), we have

$$N \cdot E[(\Delta W)_{i_1 j_1} (\Delta W)_{i_2 j_2}] \approx (W)_{i_1 i_2} (W)_{j_1 j_2} + (W)_{i_1 j_2} (W)_{i_2 j_1}. \tag{6.29}$$

This is also a generalization of (2.29).

6.2.5 The Error Covariances of Multivariate AR Spectral Estimate

Here we examine the error covariances of the consistent estimate of (6.4)

$$\hat{F}(s) = \frac{1}{2\pi} \hat{A}^{-1}(s)\hat{W}\cdot\hat{A}^{-T}(-s) \tag{6.30}$$

where $\hat{A}(s)$ is defined by replacing A_k in (6.5) by $\hat{A}_k$. Invoking a general theory of Mann & Wald (1944) mentioned by Akaike (1969 b), and noting the consistencies of $\hat{A}_k$, $\hat{W}_k$, the estimation error $\Delta F(s) = \hat{F}(s) - F(s)$ is asymptotically expressed as

$$\Delta F(s) \approx -A^{-1}(s)\Delta A(s)F(s) + \frac{1}{2\pi} A^{-1}(s)\Delta W\cdot A^{-T}(s)$$
$$-F(s)\Delta A^{T}(-s)A^{-T}(-s) \tag{6.31}$$

where

$$\Delta A(s) = -\sum_{k=1}^{m} \Delta A_k\, e^{-jks}. \tag{6.32}$$

Since from (6.26) ΔW and $\Delta A(s)$ are uncorrelated, we have

$$E[(\Delta F(s))_{i_1 j_1}(\Delta F(t))_{i_2 j_2}] = E[\frac{1}{4\pi^2} (A^{-1}(s)\Delta W\cdot A^{-T}(-s))_{i_1 j_1}$$
$$\times (A^{-1}(t)\Delta W\cdot A^{-T}(-t))_{i_2 j_2}] + E[(A^{-1}(s)\Delta A(s)F(s) + F(s)\Delta A^{T}(-s)$$
$$\cdot A^{-T}(-s))_{i_1 j_1}(A^{-1}(t)\Delta A(t)F(t) + F(t)\Delta A^{T}(-t)A^{-T}(-t))_{i_2 j_2}] . \tag{6.33}$$

Using (6.19) and (6.29), we calculate (6.33). Putting $A^{-1}(s) = C(s) = (c_{ik}(s))$, the first term of (6.33) is

$$\frac{1}{4\pi^2} E[\sum_{n,k} c_{i_1 n}(s)(\Delta W)_{nk}\, c_{j_1 k}(-s) \sum_{n',k'} c_{i_2 n'}(t)(\Delta W)_{n',k'}$$
$$\times c_{j_2 k'}(-t)] = \frac{1}{4\pi^2 N} \sum_{n,k,n',k'} c_{i_1 n}(s)c_{j_1 k}(-s)c_{i_2 n'}(t)$$
$$\times c_{j_2 k'}(-t)[(W)_{nn'}(W)_{kk'} + (W)_{nk'}(W)_{kn'}]$$

$$= \frac{1}{4\pi^2 N} [(A^{-1}(s)W\cdot A^{-T}(t))_{i_1 i_2} (A^{-1}(-s)W\cdot A^{-T}(-t))_{j_1 j_2}$$

$$+ (A^{-1}(s)W\cdot A^{-T}(-t))_{i_1 j_2} (A^{-1}(-s)W\cdot A^{-T}(t))_{j_1 i_2}]. \quad (6.34)$$

Also, one of the four terms arising from the second term of (6.33) becomes as follows;

$$E[(A^{-1}(s)\Delta A(s)F(s))_{i_1 j_1} (A^{-1}(t)\Delta A(t)F(t))_{i_2 j_2}]$$

$$= E[\sum_{n,k} c_{i_1 n}(s) \sum_i (\Delta A_i)_{nk}\, e^{-jis} f_{kj_1}(s) \sum_{n',k'} c_{i_2 n'}(t)$$

$$\times \sum_{i'} (\Delta A_{i'})_{n',k'}\, e^{-ji's}\, f_{k'j_2}(t)]$$

$$\approx \frac{1}{N} \sum_{n,k,n',k'} c_{i_1 n}(s) c_{i_2 n'}(t)(W)_{nn'} \sum_{i,i'} (\underset{\sim}{P}_{ii'})_{kk'}\, e^{-j(is+i't)}$$

$$\times f_{kj_1}(s) f_{k'j_2}(t)$$

$$= \frac{1}{N} (A^{-1}(s)W\cdot A^{-T}(t))_{i_1 i_2} (F^T(s) \sum_{i,i'} \underset{\sim}{P}_{ii'}\, e^{-j(is+i't)} F(t))_{j_1 j_2} \quad (6.35)$$

Defining a matrix $P(s,t)$ by

$$P(s,t) = \sum_{i,i'} \underset{\sim}{P}_{ii'}\, e^{-j(is+i't)} \quad (6.36)$$

, then by the same calculation as in (6.35), (6.33) eventually becomes

$$N\cdot E[(\Delta F(s))_{i_1 j_1} (\Delta F(t))_{i_2 j_2}] \approx$$

$$(A^{-1}(s)W\cdot A^{-T}(t))_{i_1 i_2} (F^T(s)P(s,t)F(t))_{j_1 j_2} + (A^{-1}(-s)W\cdot A^{-T}(-t))_{j_1 j_2}$$

$$\times (F(s)P(-s,-t)F^T(t))_{i_1 i_2} + (A^{-1}(s)W\cdot A^{-T}(-t))_{i_2 j_2} (F^T(s)P(s,-t)$$

$$\cdot F^T(t))_{j_1 i_2} + (A^{-1}(t)W\cdot A^{-T}(-s))_{i_2 j_1} (F^T(t)P(t,-s)F^T(s))_{j_2 i_1}$$

$$+ \frac{1}{4\pi^2} (A^{-1}(s)W \cdot A^{-T}(t))_{i_1 i_2} (A^{-1}(-s)W \cdot A^{-T}(-t))_{j_1 j_2}$$

$$+ \frac{1}{4\pi^2} (A^{-1}(s)W \cdot A^{-T}(-t))_{i_1 j_2} (A^{-1}(-s)W \cdot A^{-T}(t))_{j_1 i_2} . \tag{6.37}$$

It is easy to see that (6.37) reduces to (2.37) for d = 1.

Using the method of Makhoul (1978) for uni-variate case, we easily obtain a Cholesky decomposition of $\underset{\sim}{R}^{-1}$

$$\underset{\sim}{R}^{-1} = \begin{bmatrix} I & & & & \\ -A_{1\ m-1}^T & I & & 0 & \\ -A_{2\ m-1}^T & -A_{1\ m-2}^T & \cdot & & \\ \vdots & \vdots & & \cdot & \\ -A_{m-1\ m-1} & -A_{m-2\ m-2} & \cdots & & I \end{bmatrix} \cdot \begin{bmatrix} W_{m-1}^{-1} & & & \\ & W_{m-2}^{-1} & & 0 \\ & 0 & \ddots & \\ & & & W_0^{-1} \end{bmatrix} \tag{6.38}$$

$$\cdot \begin{bmatrix} I & -A_{1\ m-1} & -A_{2\ m-1} & \cdots & -A_{m-1\ m-1} \\ & I & -A_{1\ m-2} & \cdots & -A_{m-2\ m-2} \\ & & \ddots & & \vdots \\ & 0 & & \ddots & \vdots \\ & & & & I \end{bmatrix}$$

where A_{ik} ($i=1,\dots,k$; $k=1,\dots,m-1$) and W_k ($k=0,\dots,m-1$) are the theoretical coefficient and residual matrices arising from k-th order AR model fitting. From (6.38), P(s,t) becomes

$$P(s,t) = \sum_{k=0}^{m-1} e^{-j(m-k)(s+t)} A_k^T(s) W_k^{-1} A_k(t) \tag{6.39}$$

where

$$A_k(s) = I - \sum_{i=1}^{k} A_{ik}\, e^{-jis}. \tag{6.40}$$

Now we cosider the case where the fitting order M is much greater than the true order m, i.e., $M \gg m$. By putting $A_i = 0$ ($i=m+1,\dots,M$) we can regard $\{\underset{\sim}{X}_t\}$ as an M-th order AR process, so that (6.37) becomes valid after replacing m by M. But, obviously, $A_k(s) = A(s)$, $W_k - W$ $(k=m,\dots,M)$ follow so that (6.39) becomes

$$P(s,t) = \sum_{k=0}^{m-1} e^{-j(M-k)(s+t)} A_k^T(s) W_k^{-1} A_k(t) + D_{M-m}(s+t) A^T(s) W^{-1} A(t). \tag{6.41}$$

Since we assume $M \gg m$, the first term of (6.41) is much smaller than the remaining one. Neglecting this term and substituting (6.41) into (6.37), we have

$$\begin{aligned}&\left(\frac{1}{2\pi} A^{-1}(s) W \cdot A^{-T}(t)\right)_{i_1 i_2} \left(\frac{1}{2\pi} A^{-1}(-s) W \cdot A^{-T}(-t)\right)(1 + D_{M-m}(s+t) \\ &+ D_{M-m}(-s-t)) + \left(\frac{1}{2\pi} A^{-1}(s) W \cdot A^{-T}(-t)\right)_{i_1 j_2} \left(\frac{1}{2\pi} A^{-1}(t) W \cdot A^{-T}(-t)\right)_{i_2 j_1} \\ &\times (1 + D_{M-m}(s-t) + D_{M-m}(t-s)).\end{aligned} \tag{6.42}$$

Noting the property (4.29) of the Dirichlet kernel $D_N(.)$, we see that (6.42) is of order M only when $s = \pm t$. From $(\Delta F(s))_{ik} = (\Delta F(-s))_{ki}$, it is sufficient to consider the case $s = t$. In this case, (6.37) is expressed as

$$N \cdot E[(\Delta F(s))_{i_1 j_1} (\Delta F(s))_{i_2 j_2}] \approx \begin{cases} 2M\, f_{i_1 j_2}(s) f_{i_2 j_1}(s) & (s \neq 0, \pi) \\ 2M(f_{i_1 i_2}(0) f_{j_1 j_2}(0) + f_{i_1 j_2}(0) f_{i_2 j_1}(0)) & (s=0) \\ 2M(f_{i_1 i_2}(\pi) f_{j_1 j_2}(\pi) + f_{i_1 j_2}(\pi) f_{i_2 j_1}(\pi)) & (s=\pi) \end{cases} \tag{6.43}$$

The result takes the same form of that of the multivariate Blackman-Tukey procedure (Hannan 1970, page 280) and is a generalization of (2.39). However, it is obvious that actual statistical variabilities are much smaller than (6.43), since we can obtain a reasonable fitting order near m by using FPE or AIC.

6.2.6 Simulations and Concluding Remarks

To see the validities of the above results, numerical simulations were performed. We made 450 sets of data each of N = 1000 length generated by the following 2-variates third order AR process

$$\underset{\sim}{X}_t = \begin{bmatrix} 0.3 & 0 \\ 0 & -0.5 \end{bmatrix} \underset{\sim}{X}_{t-1} + \begin{bmatrix} 0.2 & 0 \\ 2.0 & 0 \end{bmatrix} \underset{\sim}{X}_{t-2} + \begin{bmatrix} 0.1 & 0 \\ 0.5 & 0 \end{bmatrix} \underset{\sim}{X}_{t-3} + \underset{\sim}{U}_t$$

with W = diag(1.0,1.0). Figs. 6.1 and 6.2 show the theoretical values. The agreements are good and justify the present asymptotic analysis.

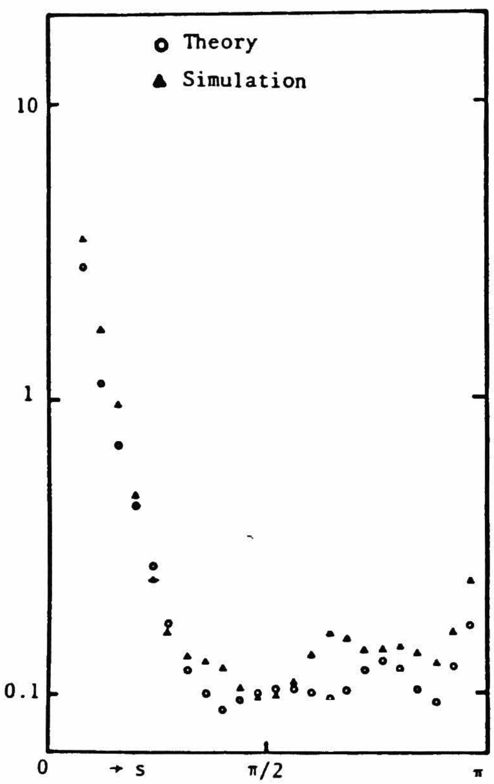

Fig. 6.1 (a)
Theoretical and experimental values of $N \cdot E[(\Delta F(s))^2_{11}]$

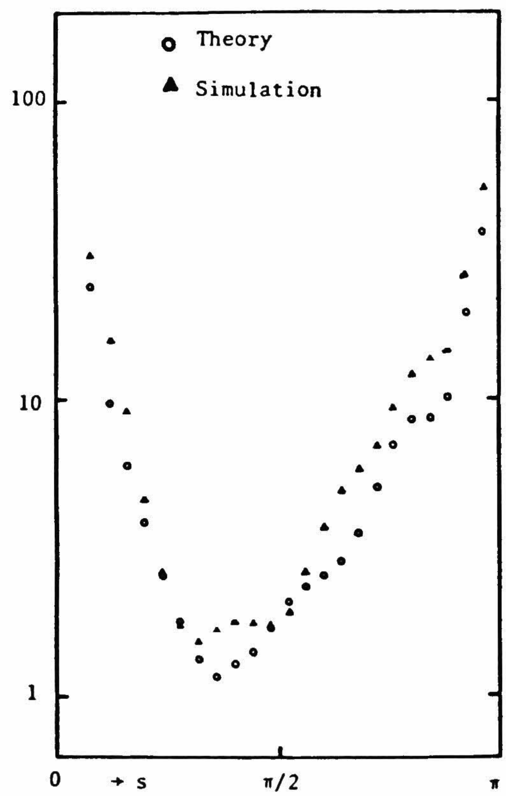

Fig. 6.2 (b)
Theoretical and experimental values of $N \cdot E[(\Delta F(s))^2_{22}]$

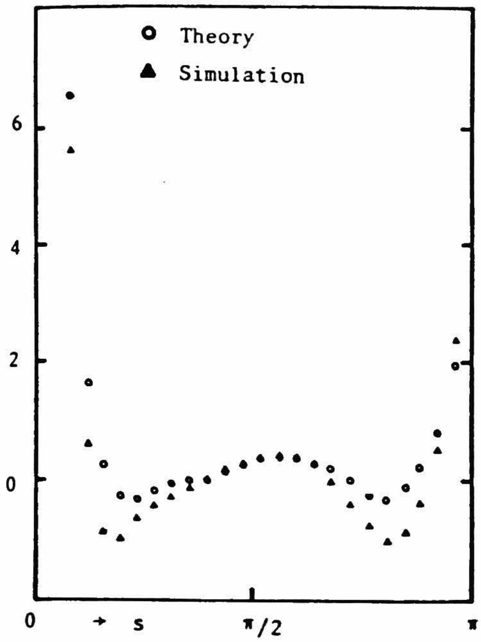

Fig. 6.2 (a)
Theoretical and experimental values of the real part of $N \cdot E[(\Delta F(s))^2_{12}]$

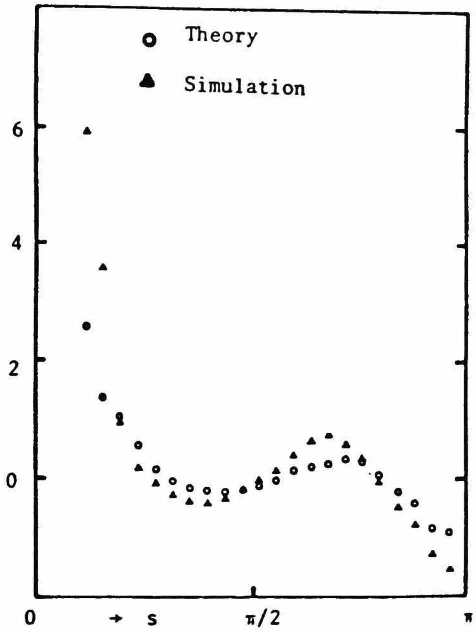

Fig. 6.2 (b)
Theoretical and experimental values of the imaginary part of $N \cdot E[(\Delta F(s))^2_{12}]$

In this section, under the assumption that $\{\underset{\sim}{X}_t\}$ is a *finite* order multivariate AR process, statistical analysis was performed and the work of Akaike (1969 b) was generalized. It is a future problem to generalize the work of Berk (1974) which treats a certain class of scalar infinite order AR processes.

6.3 Statistical Properties of Partial Autocorrelation Matrices and the Order Distribution Determined by AIC

6.3.1 Introduction

Partial autocorrelations of AR processes have many interesting properties and have been used for statistical tests (Quenouille 1947, Box & Jenkins 1970, page 65). However, as far as the author is aware , little is known about the properties of partial autocorrelation matrices of multivariate AR processes. In dealing with the multi-variate case, we must inevitably consider the backward time process which does not arise in uni-variate case.

In this section, we shall derive the asymptotic variances and co-variances of the usual forward and backward partial autocorrelation matrices and their normarlized forms due to Morf *et al.*(1978 a,b), and apply the results to determine the asymptotic distribution of the autoregression order selected by Akaike's information criterion (AIC) which can be regarded as a generalization of Shibata's result on uni-variate AR processes (Shibata 1976). Numerical examinations indicate that the probability of selecting the correct order increases as the number of variates increases.

6.3.2 Partial Autocorrelation Matrices

Consider the time series generated by (6.1). However, we do not know the true order m *a priori* so that we must fit autoregressions of various orders, say from 0 to some prefixed K to the data and choose an appropriate model. Writing (6.2) and (6.3) into one form, we have

$$[I, -A_{1k}, \dots, -A_{kk}]\ \Xi^{(k+1)} = [W_k, 0, \dots, 0] \qquad (6.44)$$

where we emphasize the dependence on the fitting order k and $\Xi^{(k)}$ is the kd×kd grand Toeplitz matrix whose (i,j)-th block matrix $\Xi^{(k)}_{ij} = R_{j-i}$. Note that $\Xi^{(m)} = \underset{\sim}{R}$. Substituting $\hat{R}_k$ in (6.7) into R_k in $\Xi^{(k)}$ yields the estimated coefficient matrices of the k-th order optimal forward prediction filter $\hat{A}_{1k},\ldots, \hat{A}_{kk}$ and the estimated prediction error covariance matrix W_k. To solve (6.44) recursively, we must accompany the backward Yule-Walker equations

$$[I, -B_{1k},\ldots, -B_{kk}]\ \Theta^{(k+1)} = [Z_k, 0,\ldots, 0] \qquad (6.45)$$

where $B_{1k},\ldots, B_{kk}$ are coefficients matrices of the k-th order optimal backward prediction filter, Z_k is the backward prediction error covariance matrix, and $\Theta^{(k)}$ is the kd×kd grand Toeplitz matrix whose (i,j)-th block matrix $\Theta^{(k)}_{i,j}$ is R_{i-j}. The details of the recursive calculations are in Whittle (1963). The estimated quantities $\hat{B}_{1k}$, ..., $\hat{B}_{kk}$ and $\hat{Z}_k$ can be obtained similarly. $\hat{A}_{kk}$ and $\hat{B}_{kk}$ are often called as (estimated) forward and backward partial autocorrelation matrices or reflection coefficient matrices.

In this section, we are most interested in their statistical properties for $k \geqq m+1$. The following well known relations hold (Morf *et al.* 1978 a,b);

$$A_{kk} = \Pi_{k-1} Z_{k-1}^{-1}, \quad B_{kk} = \Pi_{k-1}^T W_{k-1}^{-1}$$

$$\Pi_{k-1} = R_k - A_{k-1\ k-1}R_{k-1} - \cdots - A_{1\ k-1}R_1 \qquad (6.46)$$

$$W_k = (I - A_{kk}B_{kk})W_{k-1}, \quad Z_k = (I - B_{kk}A_{kk})Z_{k-1}$$

with $W_0 = Z_0 = R_0$. These relations are also valid for the estimated quantities by capping " ^ ". By using the result of Whittle (1963) and Wiggins & Robinson (1965), we obtain the following Cholesky decompositions for $(\Xi^{(k)})^{-1} = \underset{\sim}{P}^{(k)}$, $(\Theta^{(k)})^{-1} = \underset{\sim}{Q}^{(k)}$;

$$\underset{\sim}{P}^{(k)} = \begin{bmatrix} I & -B_{11}^T & \cdots & -B_{k-1\ k-1}^T \\ & I & \cdots & -B_{k-2\ k-1}^T \\ & 0 & \ddots & \vdots \\ & & & I \end{bmatrix} \cdot \begin{bmatrix} Z_0^{-1} & & & \\ & Z_1^{-1} & 0 & \\ 0 & & \ddots & \\ & & & Z_{k-1}^{-1} \end{bmatrix} \cdot \begin{bmatrix} I & & & 0 \\ -B_{11} & I & & \\ \vdots & \vdots & \ddots & \\ -B_{k-1\ k-1} & -B_{k-2\ k-1} & \cdots & I \end{bmatrix} \qquad (6.47)$$

$$
\underset{\sim}{Q}^{(k)} = \begin{bmatrix} I & -A_{11}^T & \cdots & -A_{k-1\;k-1}^T \\ & I & \cdots & -A_{k-2\;k-1}^T \\ & 0 & \ddots & \vdots \\ & & & I \end{bmatrix} \begin{bmatrix} W_0^{-1} & & & \\ & W_1^{-1} & 0 & \\ 0 & & \ddots & \\ & & & W_{k-1}^{-1} \end{bmatrix} \begin{bmatrix} I & & & \\ -A_{11} & I & 0 & \\ \vdots & \vdots & \ddots & \\ -A_{k-1\;k-1} & -A_{k-2\;k-1} & \cdots & I \end{bmatrix} . \tag{6.48}
$$

Note that in the present case, since $A_{kk} = 0$ $(k \geqq m+1)$, $\Pi_k = 0$ $(k \geqq m)$, but this implies $B_{kk} = 0$ $(k \geqq m+1)$ and $W_k = W$, $Z_k = Z$ $(k \geqq m)$. Thus we can regard $\{\underset{\sim}{X}_t\}$ as generated by the m-th order backward AR process

$$
\underset{\sim}{X}_t - B_1\underset{\sim}{X}_{t+1} - \cdots - B_m\underset{\sim}{X}_{t+m} = \underset{\sim}{V}_t \tag{6.49}
$$

with $E[\underset{\sim}{V}_t] = 0$, $E[\underset{\sim}{V}_t\underset{\sim}{V}_s^T] = \underset{\sim}{Z}\,\delta_{t,s}$ (Z : positive definite), and the stationarity condition implies

$$
|I - B_1 z - \cdots - B_m z^m| \neq 0 \quad \text{for } |z| \leqq 1. \tag{6.50}
$$

Note that the spectral density matrix $F(s)$ in (6.4) is also expressed as

$$
F(s) = \frac{1}{2\pi} B^{-1}(s) Z \cdot B^{-T}(-s) \tag{6.51}
$$

with

$$
B(s) = I - \sum_{k=1}^{m} B_k e^{jks}. \tag{6.52}
$$

6.3.3 Statistical Properties

Applying the general result of Mann & Wald (1943), the limit distribution of the elements of the estimation errors $\Delta A_{kk} = \hat{A}_{kk}$, $\Delta B_{kk} = \hat{B}_{kk}$ $(m+1 \leqq k \leqq K)$ is jointly Gaussian with a zero mean vector. In what follows, we derive the variances and covariances. By the same argument in Section 6.2.3, the estimation errors $\Delta A_{ik} = \hat{A}_{ik} - A_{ik}$, $\Delta B_{ik} = \hat{B}_{ik} - B_{ik}$ $(1 \leqq i \leqq k)$ are asymptotically expressed as

$$[\Delta A_{1k}, \dots, \Delta A_{kk}]\ \Xi^{(k)} \approx [\int_{-\pi}^{\pi} A_k(s) I_N(s) e^{js}\, ds, \dots, \int_{-\pi}^{\pi} A_k(s) I_N(s) e^{jks} ds]$$

$$[\Delta B_{1k}, \dots, \Delta B_{kk}]\ \Theta^{(k)} \approx [\int_{-\pi}^{\pi} B_k(s) I_N(s) e^{-js} ds, \dots, \int_{-\pi}^{\pi} B_k(s) I_N(s) e^{-jks} ds] \tag{6.53}$$

with (6.40) and

$$B_k(s) = I - \sum_{i=1}^{k} B_{ik}\, e^{jks}.$$

From (6.53) and noting $A_k(s) = A(s)$, $B_k(s) = B(s)$ for all $k \geqq m+1$, A_{kk} and B_{kk} $(k \geqq m+1)$ are asymptotically written as

$$\hat{A}_{kk} \approx \sum_{i=1}^{k} A(s) I_N(s) \underset{\sim}{P}_{ik}^{(k)}\, e^{jis}\, ds \tag{6.54}$$

$$\hat{B}_{kk} \approx \sum_{i=1}^{k} B(s) I_N(s) \underset{\sim}{Q}_{ik}^{(k)}\, e^{-jis} ds. \tag{6.55}$$

Hence, the asymptotic value of

$$N \cdot E[(\hat{A}_{m_1 m_1})_{i_1 j_1} (\hat{A}_{m_2 m_2})_{i_2 j_2}] \quad (m+1 \leqq m_1 \leqq m_2) \tag{6.56}$$

can be calculated by the same procedures as in (6.12)-(6.18). That is, substituting (6.54), writing it componentwise, and noting

$$\int_{-\pi}^{\pi} A(s) F(s)\, e^{jks} ds = \int_{-\pi}^{\pi} B(s) F(s)\, e^{-jks} ds = 0 \quad \text{for } k > 0 \tag{6.57}$$

, we have two terms;

$$2\pi \sum_{k,k',p,p',q,q'} \{\int_{-\pi}^{\pi} a_{i_1 p}(s) f_{pq}(s) a_{i_2 q}(-s) f_{p'q'}(-s) e^{j(k-k')s} ds$$

$$+ \int_{-\pi}^{\pi} a_{i_1 p}(s) f_{pq'}(s) a_{i_2 q}(s) f_{p'q}(-s) e^{j(k+k')s}\, ds\} (\underset{\sim}{P}_{km_1}^{(m_1)})_{p' j_1}$$

$$\times (\underset{\sim}{P}_{k'm_2}^{(m_2)})_{q' j_2} = 2\pi \sum_{k,k',p',q'} \{\int_{-\pi}^{\pi} (A(s) F(s) A^{T}(-s))_{i_1 i_2} f_{q'p'}(s)$$

$$\times e^{j(k-k')s}\, ds$$

$$+ \int_{-\pi}^{\pi} (A(s)F(s))_{i_1q'}(A(s)F(s))_{i_2p'}\, e^{j(k+k')s}ds \,\} \,(P_{km_1}^{(m_1)})_{p'j_1}$$

$$\times (P_{k'm_2}^{(m_2)})_{q'j_2}. \qquad (6.58)$$

$(1 \leqq k \leqq m_1,\ 1 \leqq k' \leqq m_2,\ 1 \leq p,p',q,q' \leq d)$

Using (6.4) and summing about p',q', we can readily show that the first part of (6.58) is equal to

$$(W)_{i_1i_2}\Big(\sum_{k,k'} P_{km_1}^{(m_1)T} R_{k-k'}^{T}\, P_{k'm_2}^{(m_2)} \Big)_{j_1j_2} \qquad (6.59)$$

From Appendix 6.1, this is simply written as

$$(W)_{i_1i_2}(Z^{-1})_{j_1j_2}\,\delta_{m_1m_2}. \qquad (6.60)$$

From Cauchy's theorem, the second part of (6.58) is zero. Thus, the asymptotic value of (6.56) is (6.60). In entirely the same way, we have the dual result

$$N\cdot E[(\hat{B}_{m_1m_1})_{i_1j_1}(\hat{B}_{m_2m_2})_{i_2j_2}] \approx (Z)_{i_1i_2}(W^{-1})_{j_1j_2}\,\delta_{m_1m_2} \qquad (6.61)$$

for $m+1 \leqq m_1 \leqq m_2$. The next quantity we are interested in is the

$$N\cdot E[(\hat{A}_{m_1m_1})_{i_1j_1}(\hat{B}_{m_2m_2})_{i_2j_2}]. \ (m+1 \leqq m_1 \leqq m_2) \qquad (6.62)$$

By the same procedure for obtaining (6.58), we have two terms;

$$2\pi \int_{-\pi}^{\pi}(A(s)F(s)B^T(-s))_{i_1i_2}\Big(\sum_{k} P_{km_1}^{(m_1)T}\, e^{jks}F^T(s) \sum_{k'} Q_{k'm_2}^{(m_2)}\, e^{jk's}\Big)_{j_1j_2} ds$$

$$+ 2\pi \int_{-\pi}^{\pi}(A(s)F(s) \sum_{k'} Q_{k'm_2}^{(m_2)}\, e^{-jk's})_{i_1j_2}(B(s)F(s) \sum_{k} P_{km_1}^{(m_1)}e^{jks})_{i_2j_1} ds.$$

$$(6.63)$$

But from (6.47) and (6.48), it can easily be shown that

$$\sum_{k} P^{(m_1)}_{\sim km_1} e^{jks} = B^T(-s)Z^{-1}e^{jm_1 s}$$

$$\sum_{k'} Q^{(m_2)}_{\sim k'm_2} e^{jk's} = A^T(s)W^{-1} e^{jm_2 s}.$$

Substituting these into (6.63) and using (6.4), (6.51) yields

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} (A(s)B^{-1}(s)W)_{i_1 i_2} (B^{-T}(s)A^T(s)W^{-1})_{j_1 j_2} e^{j(m_1+m_2)s} ds$$

$$+ \frac{1}{2\pi}\int_{-\pi}^{\pi} (I e^{-jm_1 s})_{i_1 j_2} (I e^{jm_2 s})_{i_2 j_1} ds.$$

From (6.50) and Cauchy's theorem, the first term is equal to zero and the second term is equal to

$$\delta_{i_1 j_2}\, \delta_{i_2 j_1}\, \delta_{m_1 m_2}. \tag{6.64}$$

Now we introduce the following two $d^2 \times 1$ vectors $\hat{\underset{\sim}{C}}_k$, $\hat{\underset{\sim}{D}}_k$ ($k \geqq m+1$) such that

$$(\hat{\underset{\sim}{C}}_k)_{(i-1)d+j} = \sqrt{N}\,(\hat{A}_{kk})_{ij}\ ,\ (\hat{\underset{\sim}{D}}_k)_{(j-1)d+i} = \sqrt{N}\,(B_{kk})_{ij}$$
$$(i,j=1,2,\ldots,d)$$

and the Kronecker product of two matrices A, B which is defined such that the (i,k)-th block matrix of the product $A \otimes B$ is $a_{ik}B$ where $(A)_{ik} = a_{ik}$ (Bellman 1970, page 219). Then, the results we have just obtained can be compactly stated as follows.

Proposition 6.1

The limit distribution of each $(\hat{\underset{\sim}{C}}{}_k^T\ ,\ \hat{\underset{\sim}{D}}{}_k^T)^T$ $(k = m+1, m+2, \ldots)$ *is Gaussian and independent each other with a zero mean vector and with the identical variance-covariance matrix* Φ *given by*

$$\Phi = \begin{bmatrix} \Phi_C & I_{d^2} \\ I_{d^2} & \Phi_D \end{bmatrix} \tag{6.65}$$

where $\Phi_C = W \otimes Z^{-1}$, $\Phi_D = W^{-1} \otimes Z$, *and* I_{d^2} *is the* $d^2 \times d^2$ *unit matrix.*

The proof of (6.65) is immediate from (6.60), (6.61), and (6.64). This proposition can be regarded as a mulitivariate version of the classical result of Quenouille (1947). However, it should be noted that the above distribution is *singular*, since from the algebraic property of the Kronecker products (Bellman 1970, page 236), $\Phi_C \cdot \Phi_D = (W \otimes Z^{-1})(W^{-1} \otimes Z) = (W \cdot W^{-1}) \otimes (Z^{-1} \cdot Z) \quad I_{d^2}$, that is , $\Phi_C = \Phi_D^{-1}$, thereby, the variance-covariance matrix of $\hat{\underset{\sim}{D}}_k - \Phi_C^{-1} \hat{\underset{\sim}{C}}_k$ is equal to zero matrix, indicating that the distribution is constrained on the hyper-plane determined by

$$\hat{\underset{\sim}{D}}_k = \Phi_C^{-1} \hat{\underset{\sim}{C}}_k. \tag{6.66}$$

The relation (6.66) is natural since $\hat{\underset{\sim}{C}}_k$ and $\hat{\underset{\sim}{D}}_k$ cannot vary independently because of (6.46).

Although *Proposition 6.1* states the fundamental statistical properties of the forward and backward partial autocorrelation matrices , unlike Quenouille's theorem, it explicitly contains the parameters of the process so that we cannot use it as a basis of a statistical test for the autoregression order. So let us turn our attention to the normalized autocorrelation matrices introduced by Morf *et al.* (1978 a,b). In our notation, it is expressed as

$$\rho_{k+1} = W_k^{-1/2} \, \Pi_k \, Z_k^{-T/2} \tag{6.67}$$

where $R^{-1/2}$, $R^{-T/2}$ mean $(R^{1/2})^{-1}$, $(R^{-1/2})^T$, respectively and the square root $R^{1/2}$ of a positive definite matrix R is defined as a lower triangular matrix such that

$$R^{1/2} \cdot R^{T/2} \qquad R.$$

Using this defintion and (6.46), (6.67) is written as

$$\rho_{k+1} = W^{-1/2} A_{k+1\ k+1} Z_k^{1/2}. \qquad (6.68)$$

Thus we examine the statistical properties of

$$\hat{\rho}_{k+1} = \hat{W}^{-1/2} \hat{A}_{k+1\ k+1} \hat{Z}_k^{1/2} \text{ for } k \geqq m. \qquad (6.69)$$

But $\hat{A}_{k+1\ k+1}$ is of order $O(1/\sqrt{N})$ in probability, so that (6.69) is asymptotically expressed as

$$\hat{\rho}_{k+1} \approx W^{-1/2} \hat{A}_{k+1\ k+1} Z^{1/2} \text{ for } k \geqq m \qquad (6.70)$$

, since $W_k = W$, $Z_k = Z$ for $k \geqq m$. It easily follows from *Proposition 6.1* and (6.70) that the covariance between the elements of $\hat{\rho}_k$ ($k \geqq m+1$) is

$$E[(\hat{\rho}_k)_{i_1 j_1} (\hat{\rho}_k)_{i_2 j_2}] \approx \delta_{i_1 i_2} \delta_{j_1 j_2}. \qquad (6.71)$$

Introducing an $d^2 \times 1$ vector $\hat{\underline{E}}_k$ such that $(\hat{\underline{E}}_k)_{(i-1)d+j} = \sqrt{N} (\hat{\rho}_k)_{ij}$, we obtain the following

Proposition 6.2

The limit distribution of $\{\hat{\underline{E}}_k\}$ ($k = m+1, m+2, \ldots$) is Gaussian and independent each other with a zero mean vector and the identical variance-covariance matrix I_{d^2} .

This is just an extention of Quenouille's theorem to multivariate case. *Proposition 6.2* supplies another justification of the definition of the normalized partial autocorrelation matrices.

6.3.4 The Order Distribution Determined by AIC

In fitting autoregression to the data, an important problem is to estimate the true order m from the given range $0 \leqq k \leqq K$ where we assume that $0 \leqq m \leqq K$. This problem was essentially solved by Akaike's method (Akaike 1971, 1973) by which we select the order for which the criterion

$$\mathrm{AIC}(k) = N \log |\hat{W}_k| + 2\, d^2 k \quad (0 \leqq k \leqq K) \tag{6.71}$$

takes its minimum value. We denote this order by $\hat{m}$.

Although Akaike (1971) has already obtained the statistical property of $N\cdot\log(|\hat{W}_i|/|\hat{W}_k|)$ $(m < k < i)$ which is essential for deriving the limit distribution of $\Pr(\hat{m} = k)$, we shall rederive it by using the previous results. Our method is more directly related to the original idea of Shibata (1976). From (6.46), we have

$$N\cdot\log(|\hat{W}_{m_2}|/|\hat{W}_{m_1}|) = \sum_{i=m_1+1}^{m_2} N\cdot\log|I - \hat{A}_{ii}\hat{B}_{ii}|. \quad (m+1 \leqq m_1 \leqq m_2)$$

By logarithmic expansion, *Proposition 6.1*, and (6.66), the right hand side is asymptotically equal to

$$\sum_{i=m_1+1}^{m_2} N\cdot\mathrm{tr}(\hat{A}_{ii}\hat{B}_{ii}) = \sum_{i=m_1+1}^{m_2} \hat{\underset{\sim}{C}}_i^T \hat{\underset{\sim}{D}}_i = \sum_{i=m_1+1}^{m_2} \hat{\underset{\sim}{C}}_i^T \Phi_C^{-1} \hat{\underset{\sim}{C}}_i. \tag{6.72}$$

Obviously, the limit distribution of (6.72) is χ^2 distribution with $(m_2-m_1)d^2$ degrees of freedom since $\{\hat{\underset{\sim}{C}}_i\}$ tends to a sequence of independent random variables for $i \geqq m+1$ with the same distribution $N(\underset{\sim}{0},\Phi_C)$. This result is identical with that of Akaike (1971) which was obtained by a different method. Thus the asymptotic order distribution can be obtained by entirely the same argument of Shibata (1976). It suffices to replace S_i in page 120 of his paper by $S_i = (Y_1 - 2d^2) + \dots + (Y_i - 2d^2)$ wnere $\{Y_i\}$ is a sequence of independently, identically distributed random variables $\chi^2_{d^2}$ having χ^2 distribution with d^2 degrees of freedom. Hence, the asymptotic distribution of $\hat{m}$ is given by (3.3) and (3.4) in the above paper provided that α_i is replaced by $\alpha_i = \Pr(\chi^2_{id^2} > 2id^2)$.

Table 6.1 shows some numerical values of the asymptotic order distribution for d = 2, K = 10. Table 6.2 provides the expected and experimentally observed frequency distributions of the order selected by (6.71) with K = 10 in 500 realizations each of N = 200 samples from the following 2-variates first order AR process

$$X_t = \begin{bmatrix} 0.5 & 0.5 \\ -0.5 & 0.5 \end{bmatrix} X_{t-1} + U_t, \; W = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix}$$

The agreement of the theoretical and experimental values is good.

It may also be noted that the probability of selecting the correct order, in this case $m = 1$, $\Pr(\hat{m} = m)$ increases from 0.7188 to 0.8834 as d increases from 1 to 2. Thus it is reasonable to expect that $\Pr(\hat{m} = m)$ tends to 1 as d increases indefinitely. The reason for this is that by the central limit theorem, $(\chi^2_{id^2} - id^2)/\sqrt{2id^2} \to N(0,1)$ as $d \to \infty$ so that $\alpha_i = \Pr(S_i > 0) = \Pr((\chi^2_{id^2} - id^2)/\sqrt{2id^2} > d\sqrt{i}/\sqrt{2}) \to 0$ for $i \geq 1$, but this implies that $\Pr(\hat{m} = m) = \Pr(S_1 \leq 0, \ldots, S_{K-m} \leq 0) \to 1$.

6.3.5 Concluding Remarks

We have obtained the multivariate version of Quenouille's classical result on the statistical properties of partial autocorrelations of scalar AR processes. The results were then applied to determine the asymptotic distribution of the order selected by AIC. This criterion is particularly useful when the number of the variates becomes large.

m \ $\hat{m}$	0	1	2	3	4	5	6	7	8	9	10
0	.88356	.08092	.02242	.00782	.00304	.00126	.00034	.00024	.00011	.00005	.00003
1	0.0	.88357	.08092	.02243	.00782	.00304	.00126	.00054	.00024	.00011	.00006
2	0.0	0.0	.88359	.08092	.02243	.00782	.00304	.00126	.00055	.00025	.00013
3	0.0	0.0	0.0	.88365	.08094	.02244	.00783	.00305	.00127	.00056	.00027
4	0.0	0.0	0.0	0.0	.88377	.08096	.02246	.00784	.00306	.00129	.00061
5	0.0	0.0	0.0	0.0	0.0	.88404	.08102	.02250	.00789	.00312	.00142
6	0.0	0.0	0.0	0.0	0.0	0.0	.88472	.08118	.02263	.00804	.00344
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	.88646	.08164	.02306	.00885
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	.89142	.08319	.02538
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	.90842	.09158
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

Table 6.1 The asymptotic distributions of the order selected by AIC for $d = 2$, $K = 10$.

Order $\hat{m}$	0	1	2	3	4	5	6	7	8	9	10
Expected frequency	0	442	40	12	4	2	0	0	0	0	0
Observed frequency	0	437	34	20	4	2	1	1	1	0	0

Table 6.2 The expected and observed frequencies of the order selected by AIC in 500 realizations for m = 1, d = 2.

6.4 Practical Algorithms for Pagano's New Estimation Method

6.4.1 Introduction

Recently, Pagano (1978) has found an one-to-one relationship between multivariate autoregression and scalar periodic autoregressions, and proposed an estimation method for the former model based on the latter which involves smaller number of parameters. He also derived a set of fundamental Yule-Walker (Y-W) type equations for estimating the parameters of the latter model and discussed various statistical properties of his method, but did not mention any practical computational algorithms.

In this section, we first derive a Levinson-type algorithm for solving the above mentioned Y-W equations. Then using this recursion , we show a circular lattice structure of the algorithm. This enables us to devise a Burg-type algorithm. It is shown that the stability of the estimated multivariate AR filter is ensured by the algorithm. The periodic AR orders, if unknown, are estimated by applying AIC method to each channel separately. Thus, we completely describe the practical computational aspects of Pagano's method.

Lastly, we modify the algorithm to an adaptive form of Griffiths for on-line computation and rapid adaptation for non-stationary conditions. The most salient feature of the present algorithm is that it consists of calculations of scalar quantities, thus completely avoiding matrix manipulations accompanied with usual multivariate processing methods. Since this circular lattice filter simultaneously performs whitening and orthogonalization of multivariate input samples, applications to failure detection and multichannel data compression are mentioned. Some numerical results are also shown.

6.4.2 Periodic and Multivariate Autoregressions

In this section, we review some results of Pagano (1978) needed for later discussions. Let us consider the d-variates m-th order AR process (6.1). We say that a process $\{Y_t\}$ a *periodic autoregression* of period d and order $(m_1,\dots,m_d)$ if for all integers t,

$$Y_t + \sum_{i=1}^{m_t} \alpha_t(i) Y_{t-i} = \varepsilon_t \tag{6.72}$$

where $\{\varepsilon_t\}$ is a sequence of uncorrelated variables with $E[\varepsilon_t] = 0$ and $E[\varepsilon_t^2] = \sigma_t^2$, $m_t = m_{t+d}$, $\sigma_t = \sigma_{t+d}$, and $\alpha_t(i) = \alpha_{t+d}(i)$, $i=1,\dots,m_t$. Denote the autocovariance of $\{Y_t\}$ by $R(s,t) = E[Y_s Y_t]$. From the definition of general periodically correlated random processes, we also note that

$$R(s,t) = R(s+d,t+d). \tag{6.73}$$

Then, we have

Theorem (Pagano) : If $\underset{\sim}{X}_t$ and Y_t are related by

$$(\underset{\sim}{X}_t)_i = Y_{i+d(t-1)} \tag{6.74}$$

, then $\{\underset{\sim}{X}_t\}$ is an AR process of order m with positive definite W if, and only if $\{Y_t\}$ is a periodic autoregression of period d and order $(m_1,\ldots,m_d)$ with positive $\sigma_1^2,\ldots,\sigma_d^2$ and, $m = \max_i[(m_i-i)/d]+1$ where for integer i, $[x] = i$ for $i \leq x < i+1$. The relation between various parameters in (6.1) and (6.72) are

$$-A_i = L^{-1}\cdot A_i' \qquad (6.75)$$

$$W = L^{-1}D\cdot L^{-T} \qquad (6.76)$$

with

$$L_{ki} = \alpha_k(k-i) \quad i < k \ (\alpha_k(0)=1,\ \alpha_k(i)=0 \text{ for } i<0)$$

$$(A_v')_{ki} = \alpha_k(dv+k-i) \qquad v = 1,\ldots,m$$

$$D = \mathrm{diag}(\sigma_1^2,\ldots,\sigma_d^2).$$

Thus, by estimating $\alpha_k(i)$ $(i = 1,\ldots,m_k,\ k = 1,\ldots,d)$ first, we can identify the multivariate AR model (6.1). Since L is an unit (ones on the diagonal) lower triangular, inversion of L and calculation of (6.75), (6.76) are quite easy.

The Y-W equations for $\alpha_k(i)$ are given by Theorem 2 of Pagano (1978) as

$$R(k,k-v) + \sum_{i=1}^{m_k} \alpha_k(i)R(k-i,k-v) = \delta_{v,0}\,\sigma_k^2. \quad (6.77)$$

When a set of data $\{\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_N\}$ is available, we first form a data set $\{Y_1,\ldots,Y_{Nd}\}$ by (6.74), then estimate $R(k,v)$ by

$$R_N(k,v) = \frac{1}{N}\sum_{i=1}^{p} Y_{k+di}Y_{v+di} \qquad (6.78)$$

where $p = [N-\max(k,v)/d]$, substitute (6.78) into (6.77), and finally obtain the estimate $\hat{\underset{\sim}{\alpha}}_k = (\hat{\alpha}_k(1),\ldots,\hat{\alpha}_k(m_k))^T$ for $\underset{\sim}{\alpha}_k$ $(\alpha_k(1),\ldots,\alpha_k(m_k))^T$, $k = 1,\ldots,d$. Pagano also showed that $\hat{\underset{\sim}{\alpha}}_k$, $k = 1,\ldots,d$, are consistent and asymptotically efficient estimators for the $\underset{\sim}{\alpha}_k$ and moreover, they are asymptotically independent, provided that

$\{X_t\}$ is Gaussian (Pagano 1978, Theorem 4). This implies that the parametrization in terms of the α_k allows us to perform statistical analysis for each channel separately.

Another advantage, he thought, is the significant reduction of the number of parameters, since m_k's are not necessarily such that $d \cdot m = m_k - k + 1$, $k = 1, \ldots, d$, so that we can model a multivariate AR model of order m with fewer than $d^2m + d(d+1)/2$ parameters for A_i's, W. This effect is quite significant when one of m_k's is large as compared to the remaining ones. In this case, the multivariate AR order m becomes large whereas many of the " intermidiate " order AR coefficient matrices A_i's are zero, so that direct fitting of the multivariate AR model is quite inefficient. However, in reality, we do not know m_k's *a priori*, so that at the first stage of identification we must estimate them by some method. We discuss this point in a later section.

6.4.3 A Levinson-Type Algorithm for Pagano's Y-W Equations

In this section, we derive a Levinson-type recursive algorithm for solving (6.77) with increasing $m_k = 0, 1, 2, \ldots$. Define

$$R_k(i) = \begin{bmatrix} R(k,k) & R(k-1,k) & \cdots & R(k-i,1) \\ R(k,k-1) & R(k-1,k-1) & \cdots & R(k-i,k-1) \\ \vdots & \vdots & & \vdots \\ R(k,k-i) & R(k-1,k-i) & \cdots & R(k-i,k-i) \end{bmatrix} \tag{6.79}$$

$$a_k(i) = (1 \quad \alpha_k(i,1) \ \ldots \ \alpha_k(i,i))^T$$

, then (6.77) is written as

$$R_k(i) \cdot a_k(i) = (\sigma_k^2(i) \ 0 \ \ldots \ 0)^T \tag{6.80}$$

with $m_k = i$. We also denote the explicit dependence of $\alpha_k(i)$ and σ_k^2 on m_k as $\alpha_k(m_k, i)$ and $\sigma_k^2(m_k)$, respectively. Obviously, (6.79) is segmented as

$$R_k(i) \;=\; \left[\begin{array}{c|c} R_k(i-1) & \\ \hline & \end{array}\right] \;=\; \left[\begin{array}{c|c} & \\ \hline & R_{k-1}(i-1) \end{array}\right]. \tag{6.81}$$

Also, from (6.73) we note the following cyclic property

$$R_0(i) \;=\; R_d(i). \tag{6.82}$$

As is usual in the derivations of Levinson-type algorithms (Wiggins & Robinson 1965, Burg 1975, Morf *et al.* 1977, Morf *et al.* 1978 a,b), we introduce an auxiliary vector defined by

$$\begin{aligned} b_k(i) \;&=\; (\beta_k(i,i)\ \ldots\ \beta_k(i,1)\ \ 1)^T \\ R_k(i)b_k(i) \;&=\; (0\ \ldots\ 0\ \ \tau_k^2(i))^T \end{aligned} \tag{6.83}$$

with $\beta_k(i,0) = 1$. Note that from positive definiteness of $R_k(i)$, $\tau_k^2(i) > 0$. Then from (6.80), (6.81), and (6.83), we have

$$R_k(i+1)\left[c_1\begin{bmatrix} a_k(i) \\ \\ 0 \end{bmatrix} + c_2\begin{bmatrix} 0 \\ \\ b_{k-1}(i) \end{bmatrix}\right] = c_1\begin{bmatrix} \sigma_k^2(i) \\ 0 \\ \vdots \\ 0 \\ \Delta_k(i) \end{bmatrix} + c_2\begin{bmatrix} \Delta_k^*(i) \\ 0 \\ \vdots \\ 0 \\ \tau_{k-1}^2(i) \end{bmatrix} \tag{6.84}$$

where

$$\Delta_k(i) \;=\; \sum_{n=0}^{i} R(k-n, k-i-1)\ \alpha_k(i,n) \tag{6.85}$$

$$\Delta_k^*(i) \;=\; \sum_{n=0}^{i} R(k-i-1+n, k)\ \beta_{k-1}(i,n) \tag{6.86}$$

for any constants c_1, c_2. We can easily show the so-called Burg relation (Burg 1975)

$$\Delta_k(i) \;=\; \Delta_k^*(i) \tag{6.87}$$

, since from (6.85), (6.80), and (6.81)

$$\Delta_k(i) = [0 \quad b_k^T(i)]\, R_k(i+1)[a_k^T(i) \quad 0]^T$$

$$= [a_k(i) \quad 0]\, R_k(i+1)[0 \quad b_k^T(i)]^T = \Delta_k^*(i).$$

Taking $c_1 = 1$ and $c_2 = -\Delta_k(i)/\tau_{k-1}^2(i)$ yields

$$a_k(i+1) = \begin{bmatrix} a_k(i) \\ 0 \end{bmatrix} - \frac{\Delta_k(i)}{\tau_{k-1}^2(i)} \begin{bmatrix} 0 \\ b_{k-1}(i) \end{bmatrix} \tag{6.88}$$

$$\sigma_k^2(i+1) = \sigma_k^2(i) - \frac{\Delta_k^2(i)}{\tau_{k-1}^2(i)} .$$

Similarly, taking $c_1 = -\Delta_k^*(i)/\sigma_k^2(i) = -\Delta_k(i)/\sigma_k^2(i)$ and $c_2 = 1$ yields

$$b_k(i+1) = \begin{bmatrix} 0 \\ b_{k-1}(i) \end{bmatrix} - \frac{\Delta_k(i)}{\sigma_k^2(i)} \begin{bmatrix} a_k(i) \\ 0 \end{bmatrix} \tag{6.89}$$

$$\tau_k^2(i+1) = \tau_{k-1}^2(i) - \frac{\Delta_k^2(i)}{\sigma_k^2(i)}$$

When k-1 becomes 0, from the cyclic property (6.83), the subscript k-1 = 0 must be replaced by d. Writing (6.88) and (6.89) componentwise, we have

Result 6.3 (A Levinson-type Recursive Algorithm) : The Y-W equations (6.77) can be solved by the following circular recursive algorithm.

1) Initial conditions (i = 0)

$$\sigma_k^2(0) = \tau_k^2(0) = R(k,k) \quad (k=1,\dots,d) \tag{6.90}$$

2) Order update from i to i+1

i) compute

$$\Delta_k(i) = \sum_{n=0}^{i} R(k-n,k-i-1)\, \alpha_k(i,n)$$

$$\alpha_k(i+1,i+1) = -\Delta_k(i)/\tau_{k-1}^2(i)$$
$$\beta_k(i+1,i+1) = -\Delta_k(i)/\sigma_k^2(i) \tag{6.91}$$

ii) update

$$\alpha_k(i+1,n) = \alpha_k(i,n) + \alpha_k(i+1,i+1)\beta_{k-1}(i,i+1-n)$$
$$\beta_k(i+1,n) = \beta_{k-1}(i,n) + \beta_k(i+1,i+1)\alpha_k(i,i+1-n) \tag{6.92}$$
$$(n = 1,\dots,i)$$

$$\sigma_k^2(i+1) = \sigma_k^2(i)(1 - \alpha_k(i+1,i+1)\beta_k(i+1,i+1))$$
$$\tau_k^2(i+1) = \tau_{k-1}^2(i)(1 - \alpha_k(i+1,i+1)\beta_k(i+1,i+1)) \tag{6.93}$$

where the subscript $k-1 = 0$ is replaced by d in (6.91), (6.92), and (6.93).

Remark 6.1 : $\alpha_k(i,i)$, $\beta_k(i,i)$ play the same role as the usual forward and backward reflection coefficient matrices (Burg 1975). In scalar case, they become identical and their absolute magnitude always less than 1 for stationary processes. But this does not hold for *scalar* $\alpha_k(i,i)$, $\beta_k(i,i)$. That is, from (6.91) and (6.93), we can see

$$0 < \alpha_k(i+1,i+1)\beta_k(i+1,i+1) < 1 \tag{6.94}$$

, but not necessarily $|\alpha_k(i+1,i+1)| < 1$ and $|\beta_k(i+1,i+1)| < 1$. A simple example illustrating this is given in Appendix 6.2.

Remark 6.2 : In actual processing, only a finite segment of data is available. By assuming that $Y_t = 0$ for $t \leqq 0$ or $t \geqq Nd+1$, (6.78) is written as

$$R_N(k,v) = \frac{1}{N} \sum_{i=-\infty}^{\infty} Y_{k+di} Y_{v+di}.$$

Thus, $R_N(k+d,v+d) = R_N(k,v)$. Hence, the algorithm remains valid when we replace $R(k,v)$ in (6.77), (6.90) by $R_N(k,v)$ in (6.78).

If the order $(m_1,\dots, m_d)$ is unkown, we can obtain a reasonable estimate by the well known Akaike's method (Akaike 1973). From Appendix 6.2, AIC is given by

$$\mathrm{AIC}(m_1,\dots, m_d) = \sum_{k=1}^{d} \mathrm{AIC}(m_k) \tag{6.95}$$

with

$$\mathrm{AIC}(m_k) = N\cdot\log \hat{\sigma}_k^2(m_k) + 2\, m_k$$

where $\hat{\sigma}_k^2(m_k)$ is the estimate of $\sigma_k^2(m_k)$ according to the method in *Remark 6.2*. Thus, the total minimization of (6.95) can be accomplished by minimizing each $\mathrm{AIC}(m_k)$ about m_k separately.

6.4.4 A Circular Lattice Structure of the Algorithm

Define the following i-th order k-th channel forward and backward prediction errors

$$\begin{aligned} \varepsilon(i,k+nd) &= Y_{k+nd} + \sum_{t=1}^{i} \alpha_k(i,t) Y_{k+nd-t} \\ \eta(i,k+nd) &= Y_{k+nd-i} + \sum_{t=1}^{i} \beta_k(i,i+1-t) Y_{k+nd-t+1} \end{aligned} \tag{6.96}$$

for $k = 1,\dots, d$ and integers $n \geqq 0$, respectively. Then from (6.92), we readily obtain

Result 6.4 (A Circular Lattice Structure) : (6.96) is expressed as

$$\begin{aligned} \varepsilon(i+1,k+nd) &= \varepsilon(i,k+nd) + \alpha_k(i+1,i+1)\, \eta(i,k-1+nd) \\ \eta(i+1,k+nd) &= \eta(i,k-1+nd) + \beta_k(i+1,i+1)\, \varepsilon(i,k+nd) \end{aligned} \tag{6.97}$$

with

$$\begin{aligned} \alpha_k(i+1,i+1) &= -\frac{E[\varepsilon(i,k+nd)\eta(i,k-1+nd)]}{E[\eta^2(i,k-1+nd)]} \\ \beta_k(i+1,i+1) &= -\frac{E[\varepsilon(i,k+nd)\eta(i,k-1+nd)]}{E[\varepsilon^2(i,k+nd)]} \end{aligned} \tag{6.98}$$

Remark 6.3 : Noting $\eta(i,nd) = \eta(i,d+(n-1)d)$, this circular lattice structure is illustrated in Fig. 6.3 for $d \quad 2$ where the unit delay operator z^{-1} acts on t. This is a multivariate generalization of Itakura & Saito's lattice structure for uni-variate case (Itakura & Saito 1971).

Remark 6.4 : (6.98) follows by minimizing $E[\varepsilon^2(i+1,k+nd)]$ and $E[\eta^2(i+1,k+nd)]$ about $\alpha_k(i+1,i+1)$ and $\beta_k(i+1,i+1)$, respectively. Also note that their minimization about $a_k(i)$ and $b_k(i)$ lead to (6.80) , and (6.83), respectively.

Remark 6.5 : If Y_{k+nd} is pure periodic AR, that is, $\alpha_k(i+1,i+1) \quad 0$, $\sigma_k^2(i+1) = \sigma_k^2(m_k)$ for all $i \geqq m_k$, then from (6.98), we have $\beta_k(i+1,i+1) = 0$ but from (6.93), this does not imply $\tau_k^2(i+1) = \tau_k^2(m_k)$. Thus, we cannot regard $\eta(.,.)$ in (6.96) as pure backward periodic AR. This situation is quite different from those in the usual uni- and multivariate AR processes where auxiliary quantities like $\eta(.,.)$ correspond to pure backward AR processes. However, as is well known , these definite meaning gives rise to complications in estimating partial autocorrelations, in particular, in multivariate case (Morf *et al.* 1978 a,b). Such a complication does not occur in the present case.

Remark 6.6 : From the Schwartz inequality, (6.94) holds, provided that there is no linear dependence between the components of $\underset{\sim}{X}_t$.

6.4.5 A Burg-Type Algorithm and Its Adaptive Form

Based on *Result 6.4*, we readily obtain a Burg-type algorithm which directly estimates the reflection coefficients $\alpha_k(i+1,i+1)$,

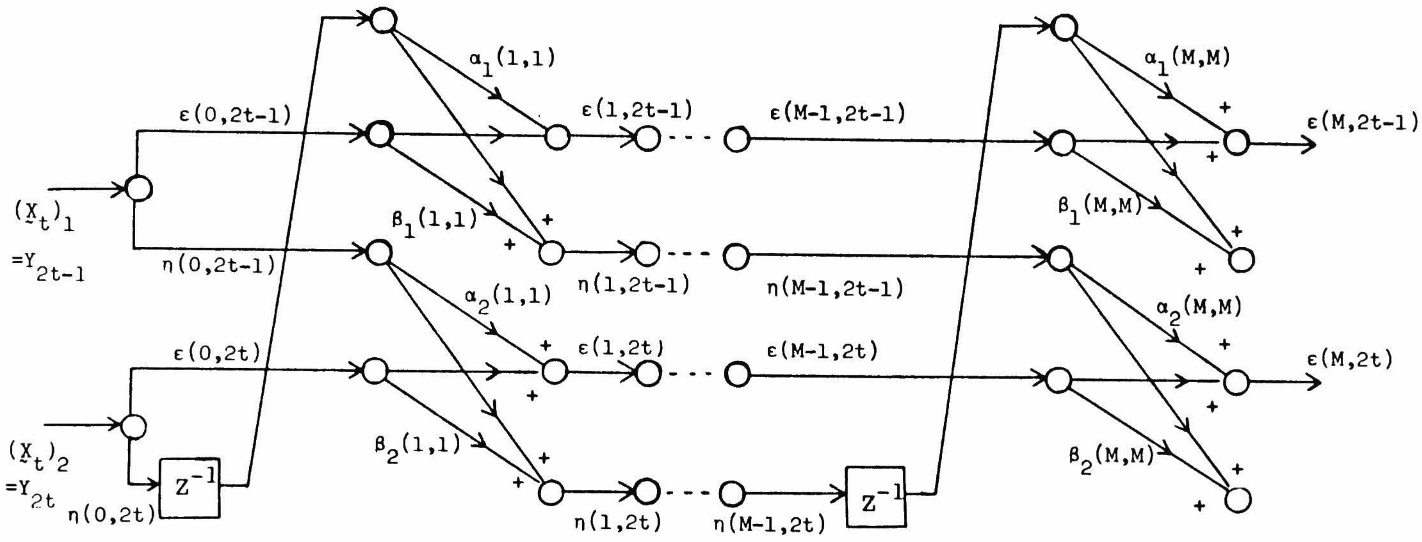

Fig. 6.3 The circular lattice structure.

$\beta_k(i+1,i+1)$ from the data.

Result 6.5 (A Burg-Type Algorithm) :

$$\hat{\alpha}_k(i+1,i+1) = -\frac{\sum_{n=L(k,i)}^{N-1} \hat{\varepsilon}(i,k+nd)\hat{\eta}(i,k-1+nd)}{\sum_{n=L(k,i)}^{N-1} \hat{\eta}^2(i,k-1+nd)}$$

$$\hat{\beta}_k(i+1,i+1) = -\frac{\sum_{n=L(k,i)}^{N-1} \hat{\varepsilon}(i,k+nd)\hat{\eta}(i,k-1+nd)}{\sum_{n=L(k,i)}^{N-1} \hat{\varepsilon}^2(i,k+nd)} \tag{6.99}$$

with $L(k,i) = [(i+1-k)/d] + 1$ give consistent estimates of $\alpha_k(i+1,i+1)$, $\beta_k(i+1,i+1)$ where $\hat{\varepsilon}(i,k+nd)$ and $\hat{\eta}(i,k-1+nd)$ are outputs of the i-th stage of the lattice filter replacing $\alpha_k(n,n)$, $\beta_k(n,n)$, $0 \leq n \leq i$, by their estimates.

Proof : The consistency property follows by an inductive argument. The upper and lower limits in the summations in (6.99) are obtained as follows. Since in $\hat{\varepsilon}(i,k+nd)$, the time indeces of the first and the last samples are $k+nd-i$ and $k+nd$, respectively. Thus, $k + \min(n)\cdot d - i > 0$ and $k + \max(n)\cdot d \leq Nd$. Similarly, for $\hat{\eta}(i,k-1+nd)$ we must have $k-1+ \min(n)\cdot d -i > 0$ and $k-1+\max(n)\cdot d \leq Nd$. The intersection of both intervals is $(i+1-k)/d < \min(n) \leq n \leq \max(n) \leq N-k/d$. Hence, $L(k,i) \leq n \leq N-1$ follows.

Remark 6.7 : From (6.99), in general, we have $0 < \hat{\alpha}_k(i+1,i+1)\hat{\beta}_k(i+1,i+1) < 1$. This guarantees the positiveness of $\sigma_k^2(i+1)$ for all k and i. This in turn assures the positive definiteness of $\hat{W}$ by (6.76). Hence, the minimum phase property of the estimated multivariate filter is followed (Morf *et al.* 1978 a).

Now we modify (6.99) to an adaptive form of Griffiths (1977, 1978) for on-line computation and rapid adaptation for non-stationary conditions. As is stated by Makhoul (1978), such an adaptive form can be obtained by replacing the arithmetic average in (6.99) by the geometric moving average. We denote the explicit dependence of $\hat{\alpha}_k(i+1,i+1)$ on N as $\hat{\alpha}_k(i+1,N)$. Then, by the same derivation with Makhoul (1978), we have

Result 6.6 (An Adaptive Circular Lattice Algorithm) : The adaptive form of (6.99) is given by

$$\hat{\alpha}_k(i+1,N+1) = \hat{\alpha}_k(i+1,N) - \frac{\nu}{\hat{G}_k(i,N+1)}\hat{\eta}(i,k-1+Nd)\hat{\varepsilon}(i+1,k+Nd)$$
$$\hat{\beta}_k(i+1,N+1) = \hat{\beta}_k(i+1,N) - \frac{\nu}{\hat{F}_k(i,N+1)}\hat{\varepsilon}(i,k+Nd)\hat{\eta}(i+1,k+Nd) \tag{6.100}$$

where

$$\hat{G}_k(i,N+1) = \mu\,\hat{G}_k(i,N) + (1-\mu)\hat{\eta}^2(i,k-1+Nd)$$
$$\hat{F}_k(i,N+1) = \mu\,\hat{F}_k(i,N) + (1-\mu)\hat{\varepsilon}^2(i,k+Nd) \tag{6.101}$$

with $0 < \mu < 1$, $\mu = 1 - \nu$. This algorithm also guarantees the fliter stability.

One of the advantages of the above two algorithms over traditional ones is that they do not involve any matrix manipulations. For example, the well known multichannel Burg-type algorithm due to Morf *et al.* (1978 a) requires the Cholesky decompositions of the mean-squared forward and backward prediction error matrices to obtain the normalized reflection coefficients in each stage of recursion as well as ordinary matrix manipulations. As is suggested in (6.76), the present method directly calculates the components of the matrix L which performs the Cholesky decomposition of W, thus avoiding such a computation in each stage of recursion. For this reason, we feel that the present method is particularly suited to adaptive processing.

Another advantage is that we can locate a change in system parameters and understand a pattern of the change much more easily, since the present parametrization serves as a system of orthogonal coordinates in the parameter space. Also, with sufficiently large M, under a steady state condition, the filter outputs $\varepsilon(M,d(t-1)+k)$, $k = 1, \ldots, d$, or the innovations are distributed independently in each channel , so that statistical tests about the innovations for detecting sudden changes or failures in the system can be performed separately in each channel.

6.4.6 Numerical Examples

To illustrate the above discussions, some numerical simulations are performed. We generate two sets of simulated two variates data of length N = 500. The reflection coefficients are estimated by the Burg-type algorithm in *Result 6.5*. In the course of the estimation , we also obtain the innovations or the outputs of the circular lattice filter. Auto and cross correlation coefficients are calculated for the innovations of length 250. Tables 6.3 and 6.4 show the results for the data

$$\begin{aligned} (\underset{\sim}{X}_t)_1 &= -a_1(\underset{\sim}{X}_t)_2 + \varepsilon_{1,t} \\ (\underset{\sim}{X}_t)_2 &\quad -a_1 a_2(\underset{\sim}{X}_{t-1})_2 + \varepsilon_{2,t} - a_2\varepsilon_{1,t} \end{aligned} \qquad (6.102)$$

with $a_1 = 1.5$, $a_2 \quad 0.5$, and $E[\varepsilon_{i,t}\varepsilon_{i,s}] = \delta_{ts}$ $(i=1,2)$, $E[\varepsilon_{1,t}\varepsilon_{2,s}] = 0$. In this case, from Appendix 6.2, we take $M \quad 1$. For comparison, these tables also contain auto and cross correlation coefficients of

$\rho_1(k)$	k = 1	2	3	4	5	6	7	8	9	10
Before Filtering	0.733	0.541	0.417	0.331	0.285	0.243	0.186	0.190	0.166	0.102
After Filtering	-0.108	0.060	-0.118	0.079	-0.052	0.073	0.030	-0.059	0.042	-0.010
$\rho_2(k)$	k = 1	2	3	4	5	6	7	8	9	10
Before Filtering	0.726	0.562	0.407	0.295	0.240	0.188	0.154	0.142	0.132	0.075
After Filtering	0.062	0.007	0.024	-0.099	0.022	0.022	-0.079	0.047	0.067	0.024

Table 6.3 The autucorrelation $\rho_1(k)$ for channel 1 and $\rho_2(k)$ for channel 2 for the data (6.102).

$\rho_{12}(k)$		k = -10	-9	-8	-7	-6	-5	-4	-3	-2	-1
Before Filtering		0.011	-0.076	-0.120	-0.159	-0.151	-0.200	-0.273	-0.298	-0.430	-0.555
After Filtering		0.080	-0.010	0.060	-0.070	0.072	0.011	-0.089	0.000	-0.081	0.094
0	1	2	3	4	5	6	7	8	9	10	11
-0.770	-0.916	-0.711	-0.538	-0.417	-0.320	-0.268	-0.225	-0.185	-0.187	-0.173	—
0.074	0.105	-0.071	-0.007	0.003	-0.084	0.005	-0.055	-0.027	-0.039	-0.107	—

Table 6.4 The cross correlation coefficients $\rho_{12}(k)$ between channel 1 and channel 2 for the data (6.102).

$\rho_1(k)$	k = 1	2	3	4	5	6	7	8	9	10
Before Filtering	0.421	0.015	-0.500	-0.617	-0.476	0.019	0.461	0.619	0.467	0.001
After Filtering	-0.068	-0.033	0.084	0.183	0.008	0.046	0.069	-0.027	0.002	0.009
$\rho_2(k)$	k = 1	2	3	4	5	6	7	8	9	10
Before Filtering	0.318	-0.028	-0.279	-0.402	-0.285	0.049	0.199	0.370	0.276	-0.001
After Filtering	0.039	-0.065	0.075	-0.069	0.030	0.038	-0.177	0.029	0.073	0.011

Table 6.5 The autocorrelation $\rho_1(k)$ and $\rho_2(k)$ for the data (6.103).

$\rho_{12}(k)$		k = -10	-9	-8	-7	-6	-5	-4	-3	-2	-1
Before Filtering		-0.023	0.317	0.486	0.362	0.088	-0.325	-0.495	-0.419	-0.059	0.321
After Filtering		0.081	-0.024	0.036	-0.044	0.102	0.044	-0.077	0.010	-0.048	0.033
0	1	2	3	4	5	6	7	8	9	10	11
0.826	0.391	0.044	-0.345	-0.464	-0.422	-0.024	0.298	0.459	0.378	0.015	—
0.012	0.038	0.033	-0.007	-0.003	0.091	0.015	-0.064	-0.034	-0.042	-0.079	—

Table 6.6 The cross correlation coefficients $\rho_{12}(k)$ for the data (6.103).

$(\underset{\sim}{X}_t)_1$ and $(\underset{\sim}{X}_t)_2$. The effects of circular lattice filtering are obvious. Similar results are obtained for the adaptive algorithm in *Result 6.6*. Tables 6.5 and 6.6 contain the results for the data generated by

$$\begin{aligned} (\underset{\sim}{X}_t)_1 &= 2\cos(0.25\pi t) + \varepsilon_{1,t} \\ (\underset{\sim}{X}_t)_2 &= 0.9\,(\underset{\sim}{X}_t)_1 + \varepsilon_{2,t} \end{aligned} \tag{6.103}$$

with M = 9.

6.4.7 Discussion and Concluding Remarks

Another application of the circular lattice filtering is the use for multichannel data compression. Fig. 6.4 depicts a diagram of the proposed system which can be regarded as a multivariate version of DPCM or ADPCM (Flanagan *et al.* 1979). As was discussed frequently, the first block performs whitening and orthogonalization of source samples so that the filter outputs $\varepsilon(M,d(t-1)+k)$, $k = 1, \ldots, d$ can be encoded *in parallel*. This greatly simplifies the encoding procedure and reduces the encoding time. The reproduction can be obtained by passing the decoded innovations through the inverse circular lattice filter whose diagram is shown in Fig. 6.5.

However, the above is only a rough sketch and further investigations are required for complete characterization of the system. In conclusion, as can be seen from the above developments, the present method offers one promising approach for various signal processing problems of multivariate time series, including adaptive and non-adaptive estimation of parameters, detection of failures, and design of efficient data compression systems. Also the circular lattice structure is quite suited to hardware implementation.

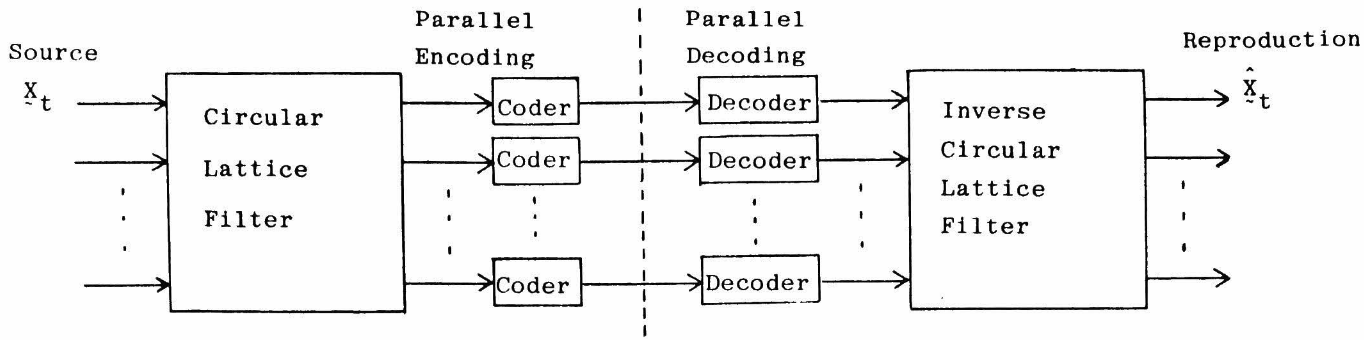

Fig. 6.4 A diagram of the proposed multichannel data compression system.

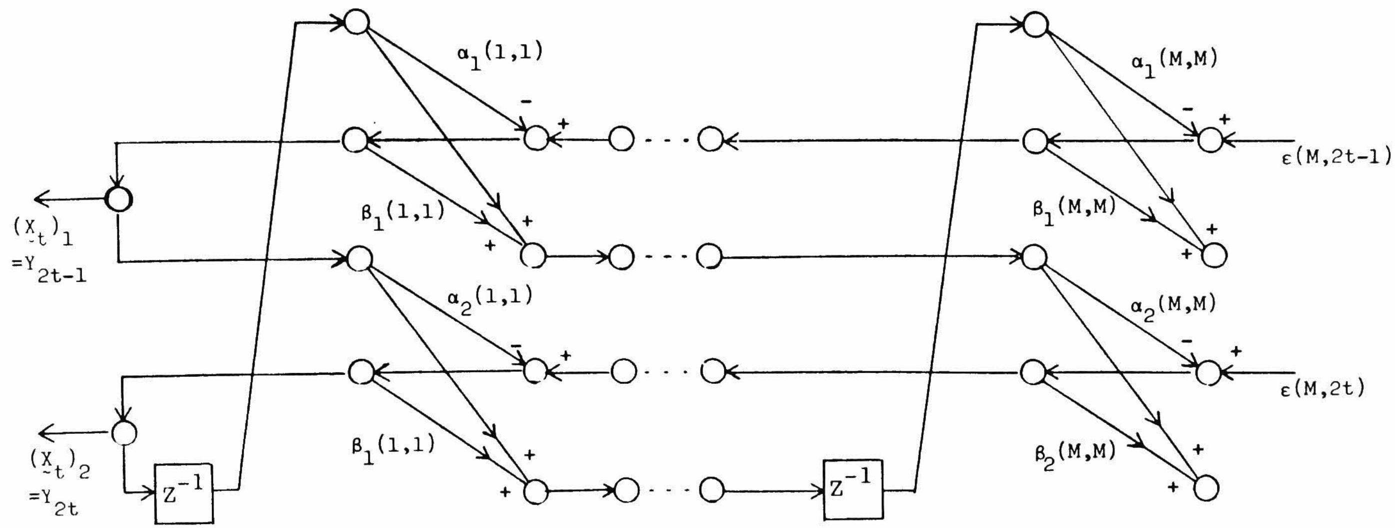

Fig. 6.5 The inverse circular lattice filter.

6.5 Conclusion

In this chapter, we have discussed various problems on multivariate AR processes including statistical properties of multivariate AR spectral estimation and partial autocorrelation matrices, distributions of the order selected by AIC, and derivation of efficient algorithms for Pagano's new estimation method. Most of the results are generalizations of those of univariate cases. But we can find some interesting things. The one thing is that the probability of selecting the correct order by AIC tends to 1 as the number of variates increases. This is a rather surprising fact since we are apt to think that increased number of parameters to be estimated causes a larger statistical variation, thereby , deteriorating the performance of the AIC statistics.

The next thing we note is that the magnitudes of the forward and backward reflection coefficients of a periodic AR process are not necessaliry less than one. In univariate case, or for a periodic AR of period 1, they happen to be equal and their magnitudes become less than one, showing the peculiarity of period 1 processes.

Though Pagano (1978) emphasizes that his new method offers statistically better estimates for multivariate AR parameters than the traditional one, regrettably, from our numerical experiments, we can not say that the former outperforms the latter. But this is reasonable since Pagano's method is optimal for estimating the parameters of periodic AR processes while the traditional method also offers the maximum likelihood estimates of multivariate AR parameters. We think that the true impact of Pagano's new idea to signal processing is that it offers a starting point for deriving the circular lattice structure in Fig. 6.3, a generalization of the celebrated Itakura &

Saito's lattice structure for the PARCOR speech analysis-synthesis system.

Appendix 6.1

Proof of (6.60)

Since $P^{(m_1)T}_{km_1} = P^{(m_1)}_{m_1 k}$, $R^T_{k-k'} = R_{k'-k}$, the second factor of (6.59) is the (j_1,j_2)-th element of the (m_1,m_2)-th block matrix of

$$(\Xi^{(m_1)})^{-1}\,[\Xi^{(m_1)} \;\; H]\begin{bmatrix} \Xi^{(m_1)} & H \\ H^T & \Xi^{(m_2-m_1)} \end{bmatrix}^{-1} \qquad \text{(A-6.1)}$$

where the last matrix is equal to the inverse of $\Xi^{(m_2)}$. But, by the well known matrix inversion lemma, (A-6.1) becomes

$$[I_{d(m_1+1)} \;\; L]\begin{bmatrix} (\Xi^{(m_1)})^{-1} + LM^{-1}L^T & -LM^{-1} \\ -M^{-1}L^T & M^{-1} \end{bmatrix} = [(\Xi^{(m_1)})^{-1} \;\; 0]$$

with $L = (\Xi^{(m_1)})^{-1}H$, $M = \Xi^{(m_2-m_1)} - H^T(\Xi^{(m_1)})^{-1}H$. Noting that $(\Xi^{(m_1)}_{m_1 m_1}) = Z^{-1}$ completes the proof. Similarly we can show that

$$\Big(\sum_{k,k'} Q^{(m_1)}_{m_1 k} R_{k-k'} Q^{(m_2)}_{k' m_2}\Big)_{j_1 j_2} = (W^{-1})_{j_1 j_2}\,\delta_{m_1 m_2}.$$

Appendix 6.2

An example concerning Remark 6.1

Consider the following simple example for $d = 2$, $m_1 = m_2 = 1$.

$$\begin{aligned} Y_{1+2n} + a_1 Y_{2n} &= \varepsilon_{1+2n} \\ Y_{2+2n} + a_2 Y_{2n+1} &= \varepsilon_{2+2n} \end{aligned} \qquad (n = 0, 1, \dots)$$

This periodic AR process corresponds to

$$\underset{\sim}{X}_t + A \cdot \underset{\sim}{X}_{t-1} = \underset{\sim}{U}_t \qquad \text{(A-6.2)}$$

with

$$A = \begin{bmatrix} 1 & 0 \\ a_2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & a_1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & a_1 \\ 0 & -a_1 a_2 \end{bmatrix}.$$

This follows from (6.75). The eigenvalues of A are 0 and $-a_1 a_2$. Thus, the condition $|a_1 a_2| < 1$ is necessary and sufficient for the stability of (A-6.2).

Appendix 6.3

Derivation of AIC($m_1, \ldots, m_d$)

The definition of AIC is

$$\text{AIC} = -2\ (\text{Maximum of log likelihood}) + 2\ (\text{Number of fitted parameters}) \qquad \text{(A-6.3)}$$

Assuming that the samples come from the Gaussian periodic AR process (6.72), we obtain the approximate likelihood as

$$\Pr(Y_1, \ldots, Y_{Nd}) \approx \Pr(\varepsilon_1, \ldots, \varepsilon_{Nd}) = \prod_{k=1}^{d} (\sqrt{2}\pi\, \sigma_k)^{-N} \exp\{\frac{-1}{2\,\sigma_k^2} \sum_{t=1}^{N} \varepsilon_{k+d(t-1)}^2\}.$$

Hence, it follows that

$$\log \Pr(Y_1, \ldots, Y_{Nd}) \approx \sum_{k=1}^{d} \{-\frac{1}{2\sigma_k^2} \sum_{t=1}^{N} \varepsilon_{k+d(t-1)}^2 - \frac{N}{2} \log 2\pi - \frac{N}{2} \log \sigma_k^2\} \qquad \text{(A-6.4)}$$

with

$$\varepsilon_{k+d(t-1)} = Y_{k+d(t-1)} + \sum_{i=1}^{m_k} \alpha_k(i) Y_{k+d(t-1)-i}\ .$$

Partial differentiating (A-6.4) by $\alpha_k(i)$ and σ_k^2 leads to (6.77) and

$$\hat{\sigma}_k^2 = \frac{1}{N} \sum_{t=1}^{N} \varepsilon_{k+d(t-1)}^2 \qquad \text{(A-6.5)}$$

, respectively. Substitution of (A-6.5) into (A-6.4) gives the maximum log likelihood. Hence, using (A-6.3) and discarding the constant terms lead to (6.95).

Appendix 6.4

Statistical Properties of $\hat{\alpha}_k(i+1,i+1)$, $\hat{\beta}_k(i+1,i+1)$ *for* $i \geqq m_k$

It can be easily seen that the estimation errors $\Delta\alpha_k(m_k,i) = \hat{\alpha}_k(m_k,i) - \alpha_k(m_k,i)$, $i=1,\ldots,m_k$, are asymptotically expressed as

$$\sum_{i=1}^{m_k} R(k-i,k-v)\,\Delta\alpha_k(m_k,i) \approx -\sum_{i=0}^{m_k} \alpha_k(m_k,i)R_N(k-i,k-v). \qquad \text{(A-6.6)}$$
$$v = 1,\ldots,m_k$$

From (6.78), the right hand side of (A-6.6) becomes

$$-\sum_{i=0}^{m_k} \alpha_k(m_k,i)\frac{1}{N}\sum_t Y_{k-i+dt}\,Y_{k-v+dt}$$

$$= -\frac{1}{N}\sum_t \varepsilon(m_k,k+dt)Y_{k-v+dt}. \qquad \text{(A-6.7)}$$

Since $\alpha_k(i,i) = 0$ for $i \geqq m_k+1$, it follows from (6.97) that $\varepsilon(i,k+dt) = \varepsilon(m_k,k+dt) \equiv \varepsilon_{k+dt}$. Thus, from (A-6.6), (A-6.7), and (6.79), we have

$$\Delta\alpha_k(i,i) \approx -\frac{1}{N}\sum_t \varepsilon_{k+dt}\sum_{j=1}^{i}(R_{k-1}^{-1}(i-1))_{ij}\,Y_{k-j+dt}. \qquad \text{(A-6.8)}$$

Hence, for $i_1 \geqq m_{k_1}+1$, $i_2 \geqq m_{k_2}+1$

$$E[\Delta\alpha_{k_1}(i_1,i_1)\,\Delta\alpha_{k_2}(i_2,i_2)] \approx$$

$$\frac{1}{N^2}\sum_{t_1,t_2,j_i,j_2} E[\varepsilon_{k_1+dt_1}\varepsilon_{k_2+dt_2}Y_{k_1-j_1+dt_1}Y_{k_2-j_2+dt_2}]$$
$$\times (R_{k_1-1}^{-1}(i_1-1))_{i_1j_1}(R_{k_2-1}^{-1}(i_2-1))_{i_2j_2}. \qquad \text{(A-6.9)}$$

From the Gaussian assumption, we have

$$E[\varepsilon_{k_1+dt_1}\,\varepsilon_{k_2+dt_2}\,Y_{k_1-j_1+dt_1}\,Y_{k_2-j_2+dt_2}] = E[\varepsilon_{k_1+dt_1}\varepsilon_{k_2+dt_2}]$$

$$\times\, E[Y_{k_1-j_1+dt_1}Y_{k_2-j_2+dt_2}] + E[\varepsilon_{k_1+dt_1}Y_{k_1-j_1+dt_1}]E[\varepsilon_{k_2+dt_2}Y_{k_2-j_2+dt_2}]$$

$$+\, E[\varepsilon_{k_1+dt_1}Y_{k_2-j_2+dt_2}]E[\varepsilon_{k_2+dt_2}Y_{k_1-j_1+dt_1}]. \qquad \text{(A-6.10)}$$

But from the definition of $\{\varepsilon_t\}$, the second term of (A-6.10) is zero. Also, the third term is nonzero if and only if $k_1+dt_1 \leqq k_2-j_2+dt_2$, $k_2+dt_2 \leqq k_1-j_1+dt_1$. But this is impossible since the above two inequalities imply $0 \leqq -j_1-j_2$, so the third term is zero. The first term becomes $\sigma^2_{k_1+dt_1}\delta_{k_1+dt_1,k_2+dt_2}R(k_1-j_1+dt_1,k_2-j_2+dt_2)$. This is nonzero if and only if $k_1 = k_2$, $t_1 = t_2$ since $k_1-k_2 = d(t_2-t_1)$ must hold with $-d+1 \leqq k_1-k_2 \leqq d-1$. Hence, (A-6.9) becomes

$$\frac{1}{N}\sigma^2_{k_1}\;\delta_{k_1k_2}\sum_{j_1,j_2}(R^{-1}_{k_1-1}(i_1-1))_{i_1j_1}R(k_1-j_1,k_2-j_2)(R^{-1}_{k_2-1}(i_2-1))_{i_2j_2}$$

for $m_{k_1}+1 \leqq i_1 \leqq i_2$. By the same argument in Appendix 6.2, we can show that

$$\sum_{j_1,j_2}(R^{-1}_{k_1-1}(i_1-1))_{i_1j_1}R(k_1-j_1,k_1-j_2)(R^{-1}_{k_1-1}(i_2-1))_{i_2j_2}$$
$$= \delta_{i_1i_2}\,(R^{-1}_{k_1-1}(i_1-1))_{i_1i_1}.$$

But $(R^{-1}_{k-1}(i-1))_{ii} = |R_{k-1}(i-2)|/|R_{k-1}(i-1)| = \tau^2_{k-1}(i-1)$, so we obtain the desired result

$$N\,E[\Delta\alpha_{k_1}(i_1,i_1)\,\Delta\alpha_{k_2}(i_2,i_2)] = \delta_{k_1k_2}\;\delta_{i_1i_2}\;\frac{\sigma^2_{k_1}}{\tau^2_{k_1-1}(i_1-1)}\,. \qquad \text{(A-6.11)}$$

It is quite tedious to derive the statistical properties of $\hat{\beta}_k(i,i)$, $i \geqq m_k+1$, by the direct calculations as above. However, from (6.91) we note that $\hat{\beta}_k(i,i) = \hat{\tau}^2_{k-1}(i-1)\hat{\alpha}_k(i,i)/\hat{\sigma}^2_k(i-1)$. But $\hat{\alpha}_k(i,i)$ is of order $N^{-1/2}$ so that $\hat{\beta}_k(i,i)$ is asymptotically expressed as

$$\hat{\beta}_k(i,i) \approx \frac{\tau_{k-1}^2(i-1)}{\sigma_k^2} \hat{\alpha}_k(i,i), \tag{A-6.12}$$

since $\sigma_k^2(i-1) = \sigma_k^2$ for $i \geqq m_k+1$. Hence, from (A-6.11) we have

$$N\, E[\Delta\beta_{k_1}(i_1,i_1)\, \Delta\beta_{k_2}(i_2,i_2)] = \delta_{k_1k_2}\, \delta_{i_1i_2} \frac{\tau_{k_1-1}^2(i_1-1)}{\sigma_{k_1}^2}. \tag{A-6.13}$$

Also from Theorem 4 of Pagano (1978), $\{\Delta\alpha_k(i,i)\}$ is asymptotically Gaussian. Thus we finally obtain that *$N\, \hat{\alpha}_k(i+1,i+1)\hat{\beta}_k(i+1,i+1)$'s ($i \geq m_k$) are asymptotically distributed as χ^2 distribution of freedom 1 and are independent for each i and k.*

This shows that the asymptotic order distribution determined by AIC for periodic AR processes is same with that for uni-variate AR processes. (See Shibata, 1976.)

References

[1] Akaike, H. (1969 a), "Fitting autoregressive models for prediction," *Ann. Inst. Statist. Math.*, 21, 243-247.

[2] ——— (1969 b), "Power spectrum estimation through autoregressive model fitting," ibid., 21, 407-419.

[3] ——— (1970), "Statistical predictor identification," ibid., 22, 203-217.

[4] ——— (1971), "Autoregressive model fitting for control," ibid., 23, 163-180.

[5] ——— (1973), "Information theory and an extension of the maximum likelihood principle," in *2nd International Symposium on Information Theory*, B.N.Petrov & F.Csáki, Eds. Budapest : Akadémia Kiado, 267-281.

[6] ——— (1976), "Canonical correlation analysis of time series and the use of an information criterion," in *System Identification : Advances and Case Studies*, D.G.Lainiotis & R.K.Mehra, Eds. New York: Academic, 27-96.

[7] ——— & Nakagawa, T. (1972), *Statistical Analysis and Control of Dynamic Systems*, (in Japanese) Tokyo: Saiensu-Sha.

[8] Alekseev, V.G. & Savitsky, Yu.A. (1973), "On spectral analysis of gaussian random processes with missing observations," *Prob. Inform. Transmiss.*, 9, 66-72.

[9] Baggeroer, A.B. (1976), "Confidence intervals for regression (MEM) spectral estimate," *IEEE Trans. Inform. Theory*, IT-22, 534-545.

[10] ——— (1978), "Sonar signal processing," in *Applications of Digital Signal Processing*, A.V.Oppenheim, Ed. Englewood Cliffs: Prentice-Hall, Chapter 6.

[11] Bellman, R. (1970), *Introduction to Matrix Analysis,* 2nd Edition, New York: McGraw-Hill.

[12] Berk, K.N. (1974), "Consistent autoregressive spectral estimate," *Ann. Statist.,* 2, 489-502.

[13] Blackman, R.B. & Tukey, J.W. (1958), *Measurement of Power Spectra from the Point of View of Communication Engineering,* New York: Dover.

[14] Bloomfield, P. (1970), "Spectral analysis with randomly missing observations," *J. R. Statist. Soc., Ser.B,* 32, 369-380.

[15] Box, G.E.P. & Jenkins, G.M. (1970), *Time Series Analysis, Forecasting and Control,* San Francisco: Holden-Day.

[16] Brillinger, D.R. (1975), *Time Series: Data Analysis and Theory,* New York: Holt, Rinehart, and Winston, Inc..

[17] Burg, J.P. (1967), "Maximum entropy power spectral analysis," presented at the 37th Annu. Int. SEG Meeting, Oklahoma City, OK, Oct. 31.

[18] ———— (1975), "Maximum entropy spectral analysis," Ph.D. dissertation, Dept. of Geophysics, Stanford Univ., Stanford, CA.

[19] Cantoni, A. & Butler, P. (1976), "Eigenvalues and eigenvectors of symmetric centrosymmetric matrices," *Linear Algebra & Its Appl.,* 13, 275-288.

[20] Flanagan, J.L., Schroeder, M.R., Atal, B.S., Crochiere, R.E., Jayant, N.S., and Tribolet, J.M. (1979), "Speech coding," *IEEE Trans. Commun.,* COM-27, 710-737.

[21] Frost, O.L. III (1976), "Power-spectrum estimation," in *Aspects of Signal Processing,* G.Tacconi, Ed., Dordrecht-Holland: D.Reidel Pub. Co., Part 1, 125-162.

[22] Gersch, W. (1970), "Estimation of the autoregressive parameters of a mixed-autoregressive-moving average time series," *IEEE Trans. Automat. Contr.*, AC-15, 583-588.

[23] Griffiths, L.J. (1977), "A continuously adaptive filter implemented as a lattice structure," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Hartford, CT, 683-686.

[24] ———— (1978), "An adaptive lattice structure for noise cancelling applications," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Tulsa, OK, 87-90.

[25] Hannan, E.J. (1970), *Multiple Time Series*, New York: Wiley.

[26] Itakura, F. & Saito, S. (1971), "Digital filtering techniques for speech analysis and synthesis," in *Proc. 7th Int. Cong. Acoust.*, Budapest, Paper 25-C-1, 261-264.

[27] Jenkins, G.M. & Watts, D.G. (1968), *Spectral Analysis and Its Applications*, San Francisco: Holden-Day.

[28] Jones, R.H. (1962), "Spectral analysis with regularly missed observations," *Ann. Math. Statist.*, 33, 455-461.

[29] ———— (1971), "Spectrum estimation with missing observations," *Ann. Inst. Statist. Math.*, 23, 387-398.

[30] Kashyap, R.L. (1970), "Maximum likelihood identification of stochastic linear systems," *IEEE Trans. Automat. Contr.*, AC-15, 25-34.

[31] Kaveh, M. (1978), "High resolution spectral estimation via rational models," presented at the Spectrum Estimation Workshop, RADC, NY.

[32] Kinkel, J.F., Perl, J., Scharf, L.L., and Stubberud, A.R. (1979), "A note on covariance-invariant filter design and autoregressive moving average spectrum analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-27, 200-202.

[33] Kromer, R.E. (1969), "Asymptotic properties of the autoregressive spectral estimator," Ph.D. dissertation, Dept. of Statistics, Stanford Univ., Stanford, CA.

[34] Lacoss, R.T. (1971), "Data adaptive spectral analysis method," *Geophysics,* 36, 661-675.

[35] Lang, S.W. (1979), "Performance of maximum entropy spectral estimators," Master Thesis, Dept. of Electrical Engineering, M.I.T., MA.

[36] Makhoul, J. (1978), "A class of all-zero lattice digital filters : properties and applications," *IEEE Trans. Acoust., Speech, Signal Processing,* ASSP-26, 304-314.

[37] Mann, H.B. & Wald, A. (1943), "On the statistical treatment of linear stochastic difference equations," *Econometrica,* 11, 173-220.

[38] —— & —— (1944), "On stochastic limit and order relationships," *Ann. Math. Statist.,* 14, 217-226.

[39] Mayne, D.Q. (1967), "A method for estimating discrete time transfer function," in *Proc. IEE 2nd U.K.A.C. Conf. : Advances in Computer Control.*

[40] Morf, M., Dickinson, B., Kailath, T., and Vieira, A. (1977), "Efficient solution of covariance equations for linear prediction," *IEEE Trans. Acoust., Speech, Signal Processing,* ASSP- 25, 429-433.

[41] ——, Vieira, A., Lee, D.T.L., and Kailath, T. (1978 a), "Recursive multichannel maximum entropy spectral estimation," *IEEE Trans. Geosci. Electron.,* GE-16, 85-94.

[42] ——, ——, and Kailath, T. (1978 b), "Covariance characterization by partial autocorrelation matrices," *Ann. Statist.,* 6, 643-648.

[43] Neave, H.R. (1970), " Extending the frequency range of spectrum estimate by the use of two data recorders," *Technometrics,* 12, 877-890.

[44] Pagano, M. (1974), " Estimation of models of autoregressive signal plus white noise," *Ann. Statist.,* 2, 99-108.

[45] ———— (1978), " Periodic and multiple autoregressions," ibid., 6, 1310-1317.

[46] Parzen, E. (1963), " On spectral analysis with missing observations," *Sankhā , Ser. A,* 25, 383-392.

[47] ———— (1967), " Time series analysis for models of signal plus white noise," in *Advanced Seminar on Spectral Analysis of Time Series,* B.Harris, Ed. New York: Wiley.

[48] ———— (1970), " Multiple time series modeling," in *Multivariate Analysis II,* P.R.Krishnaiah Ed. New York: Academic, 389-410.

[49] Pisarenko, V.F. (1973), " The retrieval of harmonics from a covariance function," *Geophys. J. R. Astr. Soc.,* 33, 347-366.

[50] Quenouille, M.H. (1947), " A large-sample test for the goodness of fit of autoregressive schemes," *J. R. Statist. Soc.,* 110, 123-129.

[51] Rife, D.C. & Boorstyn, R.R. (1974), " Single tone parameter estimation from discrete-time observations," *IEEE Trans. Inform. Theory ,* IT-20, 591-598.

[52] Rowe, I.H. (1970), " A bootstrap method for the statistical estimation of model parameters," *Int. J. Control,* 12, 721-738.

[53] Sagara, S & Wada, k. (1977), " On-line modified least-squares parameter estimation of linear discrete dynamic systems," ibid., 25, 329-343.

[54] Scheinok, P.A. (1965), " Spectral analysis with randomly missed observations: the binomial case," *Ann. Math. Statist.,*36, 971-977.

[55] Shibata, R. (1976), "Selection of the order of an autoregressive model by Akaike's information criterion," *Biometrika*, 63, 117-126.

[56] Tokumaru, H. & Takeyasu, K. (1977), "A new method for estimating the power spectral density functions of stationary ARMA processes," (in Japanese) *Trans. Soc. Instrum. Contr. Eng.*, 13, 148-153.

[57] Tong, H. (1975), "Autoregressive model fitting with noisy data by Akaike's information criterion," *IEEE Trans. Inform. Theory*, IT-21, 476-480.

[58] Walker, A.M. (1960), "Some consequences of superimposed error in time series analysis," *Biometrika*, 47, 33-43.

[59] Whittle, P. (1963), "On fitting of multivariable autoregressions and the approximate canonical facterization of a spectral density matrix," ibid., 50, 129-134.

[60] Wiggins, R.A. & Robinson, E.A. (1965), "Recursive solution to the multichannel filtering problem," *J. Geophys. Res.*, 70, 1885-1891.

[61] Yule, G.U. (1927), "On the method of investigating periodicities in disturbed series, with special reference to Wölfer's sun-spot numbers," *Phil. Trans. Roy. Soc. London, Ser. A*, 226, 267-298.

A List of the Author's Publications

[1] On the relation between fitting autoregression and periodogram with applications (with Soeda, T. and Tokumaru, H.), *Ann. Statist.*, 7, 96-107, January 1979. (relevant to Chapters 2, 4).

[2] Statistical properties of AR spectral analysis, *IEEE Trans. Acoust., Speech, Signal Processing,* ASSP-27, 402-409, August 1979. (Chapter 5)

[3] Recursive parameter estimation of an autoregressive process disturbed by white noise (with Arase, M.), *Int. J. Control,* 30, 949-966, December 1979. (Chapter 3)

[4] Statistical analysis of a spectral estimator for ARMA processes (with Tokumaru, H.), *IEEE Trans. Automat. Contr.,* AC-25, 122-124, February 1980. (Chapter 2)

[5] Statistical properties of multivariate autoregressive spectral analysis (with Kawase, T. and Tokumaru, H.), *Trans. Soc. Instrum. Contr. Eng.,* 16, October 1980, to appear.(in Japanese). (Chapter 6)

Other publications not relevant to this dissertation are

[6] A note on feedback communication with noisy side information at the receiver (with Soeda, T. and Tokumaru, H.), *IEEE Trans. Inform Theory,* IT- 23, 384-386, May 1977.

[7] Feedback transmission of binary data with automaton receiver (with ——, & ——), *IEEE Trans. Aerosp. Electron. Systs.,* AES-14, 251-257, March 1978.

[8] Optimal linear coding schemes for feedback communication with noisy side information (with ——,& ——), *IEEE Trans. Inform. Theory,* IT-24, 381-384, May 1978.

Conclusion

In this thesis, an unified method has been devised for evaluating statistical properties of various estimators of parameters of time series by using periodograms. This technique was applied to several problems including estimation of AR parameters based on noisy or missing observations, AR spectral estimation of frequencies of sinusoids in noise, and some topics in multivariate AR processes. It can be concluded that the periodogram technique is a useful and powerful tool for statistical analyses of complex estimators.

Also, the usefulness of AR method over other models such as MA or ARMA models has been emphasized by showing a possibility that a special purpose hardware can be built for analyzing multivariate time series or for multichannel data compression with LSI techniques.